High-throughput analysis of candidate imprinted genes and allele-specific gene expression

Caroline, Sophie, Frédérique Daelemans (M.D)

A thesis submitted for the degree of
Doctor of Philosophy
at University College London
2013

Clinical and Molecular Genetics Unit,
Institute of Child Health,
University College London,
London,
WC1 1EH

I, Caroline, Sophie, Frédérique Daelemans confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In diploid human organisms, the ~20,000 genes are usually functional as two active copies or alleles. Exceptionally, some genes have only one active allele while the other is silenced. Two different groups of genes fall into this minor category; the genes that exhibit random monoallelic expression (e.g. odorant receptor genes and genes coding for immunoglobulins), and those genes exhibiting monoallelic expression in a parent-of-origin specific manner, named imprinted genes. At the outset of this study in October 2006, 56 genes in humans were known to be imprinted and 98 in mice, but the total number of imprinted genes in either species was unknown. I have used high-throughput allele-specific PCR assays to screen human term placental tissue samples for new imprinted genes. Hundreds of genes were tested either because they were predicted to be imprinted or because they were candidates for which the imprinted status was simply unknown in human term placenta. My results suggested that we are reaching saturation in the number of human placentally imprinted genes. I show that *ZNF331* is imprinted in human placenta and is part of a primate lineage-specific imprinted locus showing differential methylation. My data also highlights that parental allelic specific expression is a continuum, from imprinted monoallelic expression to partial imprinting (i.e., one parental allele is slightly more (or less) expressed than the other). This continuum suggests a requirement to sequence the transcriptome of every human tissue at each different developmental stage exhaustively to assess genes for parent-of-origin specific expression and to clearly define what imprinting means. Most importantly whether 'partial' imprinting has functional significance or is just a part of the dynamic flux of gene expression. Specifically, my results call for thorough investigation of the *ZNF331* locus in human development, physiology and parent-of-origin specific diseases.

Acknowledgements

This has been a tough challenge but I have learnt so much that every minute spent on this work was worth it. I would like to thank the many people who have helped me at the different stages of this long journey that has opened my horizons so widely.

In Cambridge, a huge thank you goes to all members of Ian Dunham's lab at the Sanger for their sometimes amused support. Ian Dunham has very kindly given me a chance, trust, space, time and supervision from the very beginning and throughout and I am very grateful for that. A special thank you for Andy Mungall, Ian Sudbery, and Matt Davies for their tremendous intellectual stimulation, british humor and support. I am endebted to Matt Ritchie and Prof. Simon Tavaré for their experience and open mind. The enthusiasm for science of my sponsor, Wolf Reik, has given me endurance.

In London, Gudrun Moore has made it possible. Her tenacity was really examplary. Her attention for my emotional and educational wellbeing were crucial for this endeavour. She has my eternal gratitude. Jenny Frost, Sayeda Abu-Amero, Chris Thalasselis and Dave Monk are amazing people and I miss their company. Thank you to Phil Stanier and Andrew Stokes for their insight.

I am very honored to have met and been able to work with Prof. Charles Rodeck at UCLH. He encouraged me to register for a PhD and also made me a better doctor. I am endebted to Prof. Peebles, Prof. Jauniaux, Prof. Chitty, Mr Pandya, Miss David, Miss Peregrine, Miss Whitten. To be able to work part-time in their hospital during my PhD has enabled me to keep stamina throughout.

A very special thank you goes to the Urizar/Migeotte team for their expert help with informatic technologies, the reading of this manuscript and friendship.

Last but most importantly, I would like to thank Guillaume Smits, my so strong partner in life, for his limitless support and his everyday efforts to make our family life beautiful. Guillaume, my children, Harold and Hadrien, my friends and my whole family have been a great source of courage. I thank them for being who they are.

Table of contents

List of Figures

List of Tables

Table of abbreviations:

ART: assisted reproductive technology

BLAST: basic local alignment search tool

BLAT: basic local alignment tool

bp: base pair(s)

BWS: Beckwith-Wiedemann Syndrome

CCDS: Collaborative Consensus coding sequence

cDNA: complementary DNA

CEPH: Centre d'Etude du Polymorphisme Humain

CGI: CpG islands

Chr: chromosome

CNS: central nervous system

COBRA: combined bisulphite restriction analysis

CpG: a C (cytosine) base followed immediately by a G (guanine) base

DMD: differentially methylated domain

DMR: differentially methylated region

ds cDNA: double-stranded cDNA

eQTL: expression quantitative trait locus

gDNA: genomic deoxyribonucleic acid

ESC: embryonic stem cells

EST: expressed sequence tag

GENEVAR: gene expression variation

GWAS: genome wide association study

IAP: intracisternal A particle

ICR: imprinting control region

ICSI: intra-cytoplasmic sperm injection

IUGR: intra-uterine growth restriction

IVF: in-vitro fertilisation

LBW: low birth weight

LINE 1: long interspersed element 1

MAD: mean absolute deviation

MAF: minor allelic frequency

MALDI-TOF: matrix-assisted laser desorption/ionization time-of-flight

NCBI: National Center for Biotechnology Information

NIH: National Institutes of Health

OMIM: online Mendelian inheritance in man

PCR: polymerase chain reaction

PGC: primordial germ cell

RFLP: restriction fragments length polymorphisms

RNA: ribonucleic acid

Rpm: revolutions per minute

Seq: data obtained using next generation sequencing (RNA-Seq, ChIP-Seq)

SNP: single nucleotide polymorphism

SRS: Silver Russell syndrome

ss cDNA: single-stranded cDNA

SSCP: single strand conformation polymorphism

TNDM: transient neonatal diabetes mellitus

TSS: transcription start site

TU: transcriptional unit

UCSC: University of California, Santa Cruz

UPD: uniparental disomy

UTR: untranslated region

XCI: X-chromosome inactivation

Publications arising from this work

High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta

Caroline Daelemans, Matthew E Ritchie, Guillaume Smits, Sayeda Abu-Amero, Ian M Sudbery, Matthew S Forrest, Susana Campino, Taane G Clark, Philip Stanier, Dominic Kwiatkowski, Panos Deloukas, Emmanouil T Dermitzakis, Simon Tavaré, Gudrun E Moore, and Ian Dunham

BMC Genet. 2010; 11: 25. Published online 2010 April 19.
doi: 10.1186/1471-2156-11-25.
PMCID: PMC2871261

Data analysis issues for allele-specific expression using Illumina's GoldenGate assay

Matthew E Ritchie, Matthew S Forrest, Antigone S Dimas, Caroline Daelemans, Emmanouil T Dermitzakis, Panagiotis Deloukas, and Simon Tavaré

BMC Bioinformatics. 2010; 11: 280. Published online 2010 May 26.
doi: 10.1186/1471-2105-11-280.
PMCID: PMC2887809

Chapter 1. Introduction.

1.1. Background to the project.

In each cell of our body there is one copy of our mother's genome and one copy of our father's genome. Both genomes are indispensable for the normal development of the embryo (BARTON *et al.* 1984; SURANI *et al.* 1984; MCGRATH and SOLTER 1986). Such dependence is due to a small subset of genes being expressed in a parent-of-origin specific manner, so-called imprinted genes. Imprinted genes, because of their extreme mode of regulation, can be medically important causing severe although rare syndromes or tumour growth if disrupted (FEINBERG 2007). Mild disruption could be involved in the physiopathology of more common diseases as, for example, Type 1 Diabetes (WALLACE *et al.* 2010), Type 2 Diabetes (YASUDA *et al.* 2008; KONG *et al.* 2009) and breast cancer (KONG *et al.* 2009).

At the start of this project, new allele-specific PCR technologies, the sequence of the human genome (LANDER *et al.* 2001; VENTER *et al.* 2001), and the HapMap catalogue of human sequence variation (THE INTERNATIONAL HAPMAP 2003) enabled researchers to study allelic expression differences on a larger scale. It was also speculated that all imprinted genes were not yet known (BARLOW 1995). Hence, we designed a strategy to thoroughly test the mode of allelic expression of hundreds of imprinted genes candidates in human term placenta.
In this thesis, I will describe imprinting of potential new candidates and the different modes of allelic expression bias and discuss their possible medical relevance. I will describe the historical screening methods and review the prediction algorithms used to identify new candidate genes.

1.2. Genomic imprinting.

1.2.1. Definition of imprinting.

A gene is imprinted when the paternally or maternally inherited allele is predominantly expressed (WALTER and PAULSEN 2003; WOOD and OAKEY 2006; FROST and MOORE 2010). The term 'imprint' was first used by Helen Crouse in 1960 when she observed, in *Sciarid* flies, that embryos, initially triploid for the X chromosome, inactivate one copy (in female embryos) or both copies (in male embryos) inherited from the father. She proposed that 'the chromosome which passes through the male germ line acquires an imprint that results in behaviour exactly opposite to the imprint conferred on the same chromosome by the female germ line' (CROUSE 1960).

Later, the gametic imprinting research field being already well developed, Denise Barlow refined the gametic imprinting definition as 'a reversible process whereby a gamete-specific modification in the parental generation can sometimes lead to functional differences between maternal and paternal genome in diploid cells of the offspring' (BARLOW 1994).

1.2.2. Discovery of imprinting.

1.2.2.1. In human.

Complete hydatiform moles are the product of an abnormal conception. Their villi are hydropic, their trophoblast is markedly hypertrophic, there is no embryo, and they have the potential to become malignant (JACOBS *et al.* 1980). The karyotype of complete hydatiform moles is usually 46,XX but they can also be either 46XY or 46YY. Kajii et al. showed by studying centromeric markers that complete hydatiform moles possessed only paternal chromosomes (KAJII and OHAMA 1977). Later, it was shown that the most likely mechanism is an empty egg fertilised by a haploid sperm that will duplicate to restore diploidy (JACOBS *et al.* 1980). In rare cases, the chromosomal make up of the mole is 46,

XY and it results from the fertilisation of an empty egg by two haploid spermatozoa (OHAMA *et al.* 1981). In these abnormal conceptuses, two functional paternal genomes are present but they originate from the same parent. A conclusion was that one-parent diploidism is not sufficient for the development of a normal placenta and a normal embryo. The opposite situation is encountered in ovarian teratomas. These benign tumours contain all embryonic tissue layers and contain only maternal genomic material. Hence, 'paternal' diploidism creates exclusively placental tissue while 'maternal' diploidism creates exclusively embryonic tissue. These observations suggested that both genomes were required for the development of a normal mammalian conceptus. During the following years, imprinting was progressively dissected thanks to naturally occurring uniparental disomies (HENRY *et al.* 1989; NICHOLLS *et al.* 1989; HENRY *et al.* 1991).

1.2.2.2. In Mice.

The first evidence of imprinted genes came from nuclear transplantation experiments in mouse embryos that showed that the paternal and the maternal genomes were not equivalent and that both were necessary for the successful development of a trophoblast and an embryo (MCGRATH and SOLTER 1983; BARTON *et al.* 1984; MANN and LOVELL-BADGE 1984; MCGRATH and SOLTER 1984; SURANI *et al.* 1984). In 1985, mice produced with uniparental duplications of entire or part of chromosomes helped to define the regions that were involved in non-complementation (CATTANACH and KIRK 1985). Disomy 11 mice, for example, are smaller when their chromosomes are of maternal origin (ratio of weight 0.6) and larger when of paternal origin (ratio of weight 1.4) (CATTANACH and KIRK 1985). Opposite phenotypes depending on the parent-of-origin of the duplication were also observed for behaviour (hypokinetic and hyperkinetic) and it was suggested that this phenomenon was due to increased versus decreased gene activity (CATTANACH and KIRK 1985).

The first imprinted genes to be discovered were the Insulin-like growth factor (*Igf2)* (DECHIARA *et al.* 1991; FERGUSON-SMITH *et al.* 1991), which is paternally

expressed, its receptor *(Igf2-r)* (BARLOW *et al.* 1991) and H19 (an expressed RNA of unknown function that lacks an open reading frame (ORF)) (BARTOLOMEI *et al.* 1991) both maternally expressed. Further imprinted genes have been discovered over the last two decades using many different methods: targeted deletions, allele-specific activity, uniparental duplications from translocation intercrosses, positional cloning, systematic examination of allelic expression of neighbouring genes within a locus (ZWART *et al.* 2001), genomic methylation screens (SMITH *et al.* 2003), and more. In 2006 at the start of the work described in this thesis, around ninety imprinted genes had been found, and the question of the completeness of the imprinted gene list was still open.

1.2.3. Evolution of imprinting.

Imprinting evolved ~ 166 million years ago after divergence from the egg-laying monotremes (KILLIAN *et al.* 2000; MURPHY and JIRTLE 2003; BININDA-EMONDS *et al.* 2007; HORE *et al.* 2007; RENFREE *et al.* 2009) (Figure 1, see next page). Several loci have been studied in depth (HORE *et al.* 2007). Imprinting of *Igf2* and its receptor *Igf2r* is only known in placental mammals (KILLIAN *et al.* 2000; O'NEILL *et al.* 2000; SUZUKI *et al.* 2005). There is no evidence of imprinting of *Igf2* or *Igf2r* in egg-laying mammals (KILLIAN *et al.* 2001; NOLAN *et al.* 2001), although prenatal tissue samples have not been studied to confirm this.

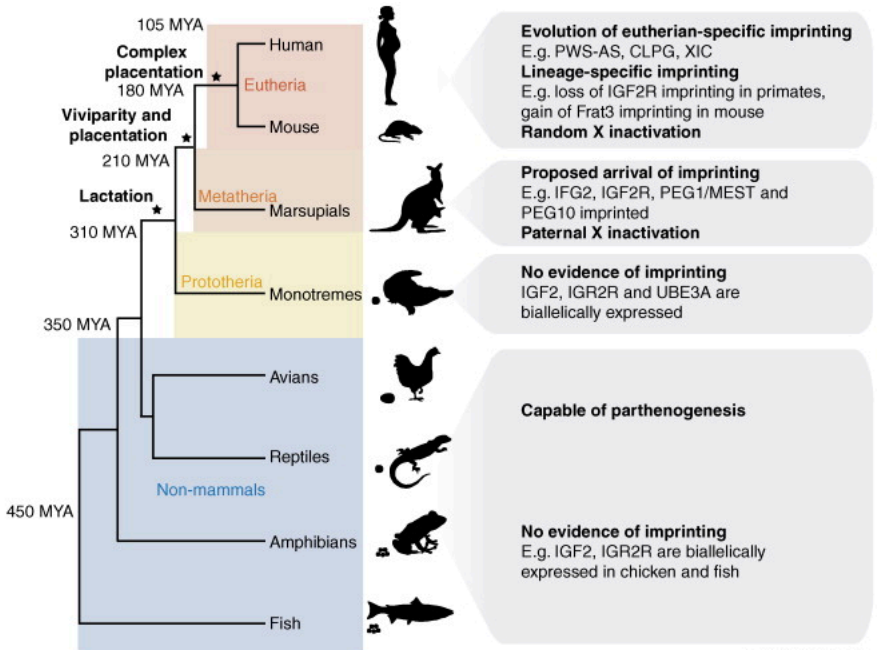

Figure 1: Evolution of imprinting.
Summary of the different loci that have been studied and are reviewed in Hore et al. (HORE *et al.* 2007). Figure reused from (HORE *et al.* 2007) with permission obtained via http://www.copyright.com/.

Several hypotheses have been proposed to explain the occurrence of imprinting (HURST 1997): the kinship theory (MOORE and HAIG 1991), the ovarian time bomb (VARMUZA and MANN 1994), the host defense model (BESTOR 1998) and, the rheostat model (BEAUDET and JIANG 2002). These hypotheses and the evidence for them are reviewed below.

1.2.3.1. The kinship theory.

The most developed hypothesis is the 'parental conflict' theory (MOORE and HAIG 1991; WILKINS and HAIG 2003) that has been more recently termed 'kinship' theory (HAIG 2004). In summary, it states that genes expressed in the offspring from paternal alleles promote prenatal and postnatal growth, while in contrast, genes expressed from maternal alleles suppress growth. In the prenatal

period, this evolutionary choice is logical if we accept the following postulate: the main evolutionary goal of reproduction is passing genes to the next generation. In mammals, the majority of species are polyandrous and polygamous. Males can have different mothers for each of their progeny. They will tend to father the strongest possible fetus/placenta/child at the cost of the mother and of the other progeny from that mother. A female *a contrario* has to spare herself at each pregnancy to ensure her survival, thereby increasing her chance to produce the greatest number of babies possible. At a molecular level, fathers will tend to overexpress growth-promoting genes and silence growth-repressing genes in the fetus and placenta, while mothers will try to do the reverse. The fight will take place in each cell of the conceptus (placenta and fetus). In the postnatal period, lactating mothers will still provide all food resources and will be manipulated in the same way by the progeny during lactation (paternal enhancement and maternal repression of demand). For some species, in the post-lactation stage, bi-parental care for baby nutrition exists and could also be manipulated by the progeny. Imprinting of certain genes could then disappear or even reverse (UBEDA 2008). With age, the progeny becomes independent for its feeding and imprinting will often disappear. Indeed imprinting brings a higher risk of haploinsufficiency and a disadvantageous level of gene expression after the fetal/neonatal period. Imprinting could also mean greater allele specific effects (ASE), the other allele being not expressed. For example, if one allele works at 80% and the other at 100%, in a biallelic mode the mean expression will be of 90%; in an imprinted mode, expression will be of 80 or 100%, which brings more extreme expression quantitative trait locus (eQTL) effects.

The kinship theory seems to be the most robust theory to explain the *appearance* of imprinting in mammals. However, some authors argue that the regulation of some genes does not fit with the kinship theory. It may be that other evolutionary pressures (see below) have used the existing imprinting mechanisms, created under the kinship theory selective pressures, to regulate tissue specific expression of other genes.

1.2.3.2. Ovarian time-bomb hypothesis.

Imprinting could act as protection against trophoblastic disease (VARMUZA and MANN 1994). In Mammals, the embryo invades the uterine epithelium aggressively compared to other phyla. Varmuza and Mann proposed that imprinting is the critical mechanism that enables mothers to defend against malignant invasion of their tissues by the trophoblast.

One argument in favour of this theory is the larger number of maternally controlled imprinted genes (REIK and WALTER 2001). A caveat is that it implies that males have *no evolutionary* advantage to imprinting (this postulate cannot be 'tested'). One Darwinian axiom can be used: nature tests everything and selection keeps the best, and this law is the same for both sexes. For the ovarian time bomb hypothesis, we need imprinting to evolve in females, but not in males. This theory is therefore less parsimonious than the kinship theory and thus less likely to explain the *appearance* of imprinting. However, it is true that the ovarian time bomb hypothesis could be sufficient to explain later acquisition of maternal imprinting in *some* genes.

1.2.3.3. The host-defense model.

This theory states that methylation of CpG islands (as in differentially methylated regions (DMRs)) is an adaptation mechanism to defend the genome against transposons (BESTOR 1998). This theory is limited and is likely to represent a useful 'molecular regulation tool acquisition' that could add itself to the kinship theory. Indeed, for imprinting to appear we need acquisition in time of *both* 'an evolutionary interest' (for selection to actively select and keep this complex epigenetic regulation) and 'the molecular tools' allowing establishment and maintenance of imprinting (SMITS *et al.* 2008). Transposons do not exist in every imprinted locus but molecular tools, important to establish imprinting, are crucial to maintain silencing of transposons during meiosis (BOURC'HIS and BESTOR 2006). Hence, it seems more logical to think that the genome has used

the same molecular tools to both control transposons and to create imprinting than to believe imprinting is a by-product of the host defence against transposons.

1.2.3.4. The rheostat model.

Beaudet and Jiang have proposed that genomic imprinting is a mechanism that maximises the interindividual variability in the levels of gene expression for dosage-sensitive loci (BEAUDET and JIANG 2002). They hypothesised that imprinted loci have a haploid selective advantage and may be variable along a continuum for their level of expression and for the resultant phenotype. This theory uses exactly the same argument than the kinship theory but is formulated in a more ASE centric manner (see section maximisation of ASE effects by imprinting).

In conclusion, the kinship theory best explains the selection of imprinting in placentas. From this moment, the epigenetic toolbox allowing imprinting became more and more diverse and different evolutionary selection mechanisms made use of it to achieve imprinting at new loci in different tissues, in different pathways (growth, behaviour) or for different goals (e.g. control and regulation of transposons).

1.2.4. Common features of imprinted genes.

1.2.4.1. Genomic clustering of imprinted genes.

Another common feature of imprinted genes is that they tend to cluster in the genome (THORVALDSEN and BARTOLOMEI 2007). In a cluster, genes can be paternally, maternally or biallelically expressed; protein-coding transcripts co-exist with non-coding RNAs, antisense RNAs, and small RNAs (REIK and WALTER 2001); some genes share regulatory mechanisms (e.g. DMRs, histone

modifications, enhancers, insulators). At each locus, there is a germline DMR, an element that controls the parental specific expression of imprinted genes across the domain. Its role as an imprinting control region (ICR) is proven when disruption of imprinting follows their targeted deletion in mouse or by microdeletions in patients. Secondary DMRs can also be present. These are methylated after fertilization (while ICRs are methylated in the germline) and also called somatic DMRs. DNA methylation has been shown to repress expression of non-coding RNAs.

1.2.4.2. DNA methylation.

DNA methylation is the covalent binding of a methyl group to cytosine nucleotides. Typically, it occurs in a CpG (cytosine-phosphate-guanine in a linear DNA sequence) dinucleotide context in somatic tissues. However, non-CpG methylation has been described and seems extensive in embryonic stem cells (ESC) (LISTER *et al.* 2009).

DNA methylation is associated with X-chromosome inactivation (XCI) (YEN *et al.* 1984), embryogenesis (LI *et al.* 1992; REIK 2007), imprinting (SAPIENZA *et al.* 1987; SWAIN *et al.* 1987; CHAILLET *et al.* 1991; SASAKI *et al.* 1991; FERGUSON-SMITH *et al.* 1993; LI *et al.* 1993; RAZIN and CEDAR 1994; BARLOW 1995; SMRZKA *et al.* 1995; LIPPMAN and MARTIENSSEN 2004) and tumorigenesis (MCKENNA and ROBERTS 2009).

In the life and development of an organism, several methylation cycles occur. These are outlined hereunder. At the cell level we must separate germ and and somatic cells. In addition, we must separate imprinted genes from all other genes.

1.2.4.2.1. *Methylation cycle in germ cells.*

Germ cells specialise very soon in embryonic life and have a special methylation cycle destined to allow reproduction (Figure 2). In germ cells, DNA methylation at every imprinted differentially methylated region (DMR) (which is differentially methylated on the maternal and paternal chromosome like in any somatic cell) is erased by mouse embryonic day 11.5 when they reach the gonadal ridges (SZABO and MANN 1995; HAJKOVA *et al.* 2002; LEE *et al.* 2002) (Figures 2 and 3). Then during germ cell maturation ICRs are methylated according to the sex of the embryo (Figures 2 and 3). In male embryos, the paternal DMR methylation is established in the fetal gonads, i.e. very early in the development (DAVIS *et al.* 2000; HAJKOVA *et al.* 2008) and completed before birth. In females, the maternal DMR methylation profile will be established postnatally at each ovulation during the development of the dominant follicle (OBATA *et al.* 1998; LUCIFERO *et al.* 2002; SCHAEFER *et al.* 2007) (Figure 2).

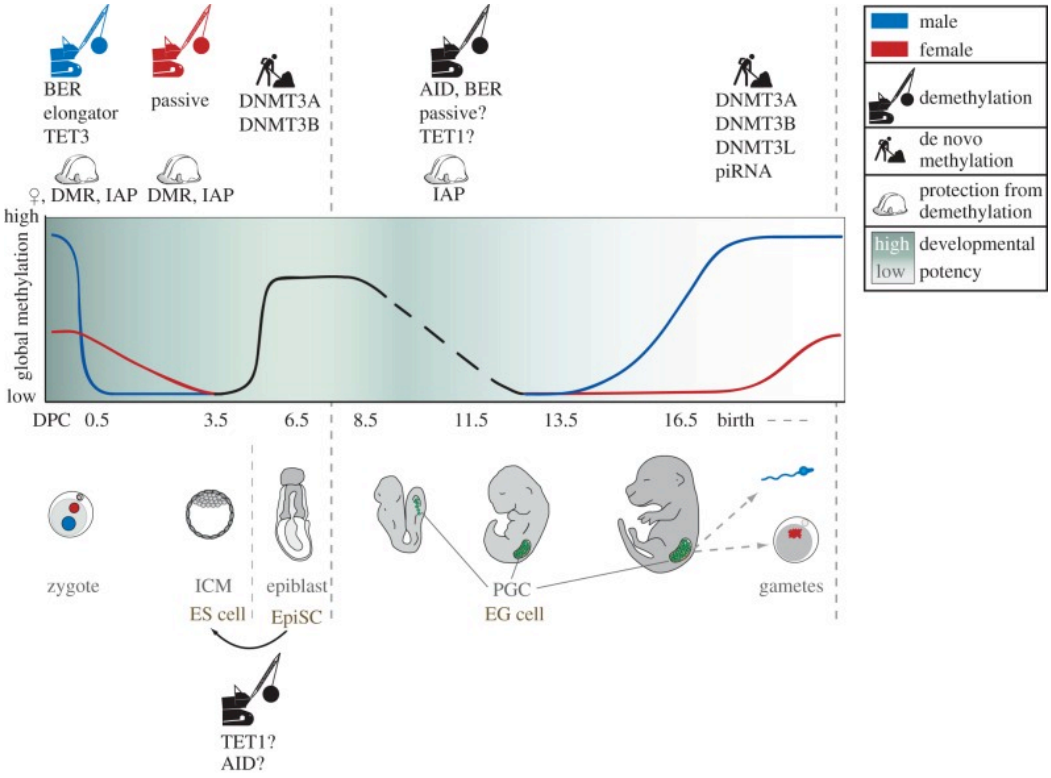

Figure 2: Summary of the methylation cycle of germ cells through life.
After fertilisation, the paternal genome is actively demethylated. At the blastocyst stage, the whole genome is passively demethylated except at a number of CpG islands and repetitive sequences (SMALLWOOD *et al.* 2011; SEISENBERGER *et al.* 2012). At midgestation, the parental imprint marks are erased in the germline to reflect the sex of the fetus. ICR stands for imprinting control region (all DMRs are candidate ICRs but need to be confirmed by the *in vivo* characterization of the DMR deletion) (SPAHN and BARLOW 2003). The picture is reused with permission from Seisenberger et al. (SEISENBERGER *et al.* 2013a)(under the http://creativecommons.org).

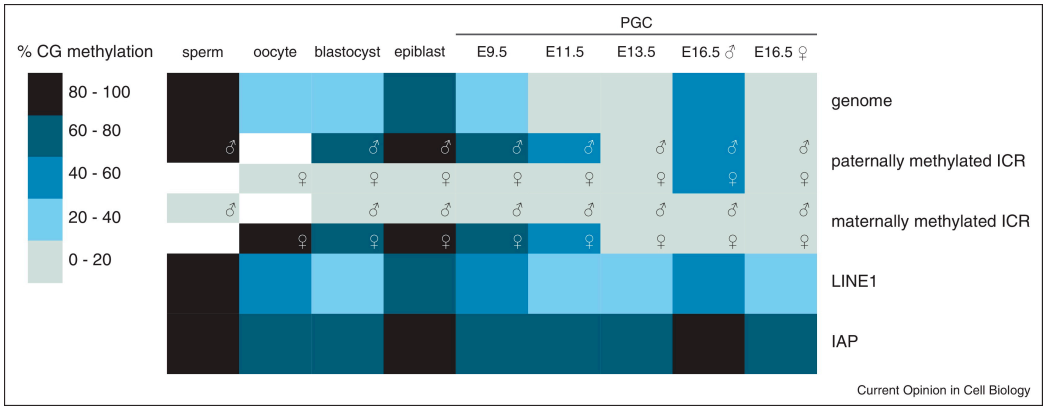

Figure 3: Methylation heatmap of the genome and several of its specific regions during mouse embryonic development.

The sperm is hypermethylated while the oocyte is relatively hypomethylated. After fertilisation, the paternal genome is actively demethylated. At implantation, active methylation occurs. Another wave of demethylation takes place in primordial germ cells (PGCs). Remethylation will take place around the embryonic day 14.5 in male cells only. It occurs in oocytes postnatally. Repetitive elements like the intracisternal A particles (IAP) remain highly methylated throughout embryonic development (SMITH *et al.* 2012). Figure reused with permission from Seisenberger et al. (SEISENBERGER *et al.* 2013b).

For a paternally methylated DMR, as the *IGF2-H19* locus (Figure 5), the *H19* DMR (which controls the whole *IGF2-H19* locus) is unmethylated on the maternal chromosome and methylated on the paternal one in somatic cells and very early germ cells. In the fetal gonad, all imprinted DMR methylation will be reset to zero. If the embryo is a male, quite rapidly the testis specific CTCFL (also known as BORIS) will bind the *H19* DMR and start an epigenetic process finally resulting in DNA methylation of the DMR on both chromosomes (BELL and FELSENFELD 2000; HARK *et al.* 2000; SZABO *et al.* 2000; SCHOENHERR *et al.* 2003; JELINIC *et al.* 2006). Hence, post meiosis, every spermatozoid will have one chromosome bearing a methylated *H19* DMR, which corresponds to the expected paternal epigenotype. If the embryo is a female, the oocyte DMR will bind CTCF plus several other epigenetic proteins (BELL and FELSENFELD 2000; HARK *et al.* 2000; SZABO *et al.* 2000; SCHOENHERR *et al.* 2003; (FEDORIW *et al.* 2004; ENGEL *et al.* 2006) and stay unmethylated. So whatever the chromosome retained during meiosis, the *H19* DMR present on the maternal chromosome will always be unmethylated (the expected epigenotype for a maternal chromosome).

It has been shown that the methyltransferases DNMT3A, 3B and 3L are crucial to germ cell imprinted DMR methylation cycle (OKANO *et al.* 1999; BESTOR 2000; BOURC'HIS *et al.* 2001; CHEDIN *et al.* 2002; HATA *et al.* 2002; ARNAUD *et al.* 2006) (Figure 2). DNMT3A is a *de novo* methyltransferase (KANEDA *et al.* 2004) and interacts with DNMT3L (JIA *et al.* 2007), its regulatory factor. DNMT1 is responsible for the maintenance of methylation (CHEN and LI 2004). Interestingly, those proteins seem to appear after the divergence of protherians and therians in mammalian evolution suggesting that imprinting - as we know it in human or mouse - could not exist in the monotremes (platypus, echidna) germ cells (YOKOMINE *et al.* 2006; HORE *et al.* 2008; SMITS *et al.* 2008).

Recently, it has been shown that sex-specific methylation was not restricted to imprinted loci (SMALLWOOD *et al.* 2011; KOBAYASHI *et al.* 2012; SMITH *et al.* 2012). The first genome-wide study revealed that approximately 900 CpG islands were specifically methylated in mature oocytes and only 60 in mature sperm (SMALLWOOD *et al.* 2011). The sex-specific methylated CGIs include the imprinted gDMRs and many more.

After fertilisation, there is an epigenetic reprogramming that affects the male and female genome in a different manner. The male genome is actively and rapidly demethylated. The mechanism is not fully elucidated but the role of Tet proteins has been highlighted (GU *et al.* 2011). The female genome is slower to demethylate (Figure 2): the methylation is not maintained at each cell division and is passively progressively lost. After fertilisation, methylation of ICRs are not erased and maintain the mark of their parental origin.

1.2.4.2.2. *Methylation cycle in somatic cells.*

Each cell will have a tendency to shut down the genes that are not necessary for its own function with the exception of the genes that are required for the function of its "daughters" (REIK 2007). For example, an intestine stem cell will express genes corresponding to 'pluripotency' (capacity of asymetric continuous cell division – one daughter cell staying the pluripotent progenitor, the other one becoming the more specialised cell) and will ensure that the genes necessary for

the correct functioning of her specialised daughter cell are demethylated (microvilli cell).

1.2.4.2.3. *Medical relevance of cellular methylation cycles.*

Throughout development, there is a need for proteins to establish and maintain the methylation (or hypomethylation) located at DMRs. The dysregulation of these methylation proteins is the cause of several diseases. For example, ICF syndrome (immunodeficiency, centromere instability and facial abnormalities, MIM 242860) patients have a decreased level of immunoglobulins and die from infection. It is a genomic methylation disorder due to mutations in the DNA methyltransferase gene 3B, a gene necessary for the establishment and maintenance of methylation at germinal DMRs and gene promoters (XU *et al.* 1999; JIN *et al.* 2008). Also in humans, a methylation defect causes a rare condition called biparental complete hydatiform mole (MIM 231090 and 614293). As outlined previously 'classical non recurrent' hydatiform moles are usually androgenetic in their make-up. In recurrent familial biparental moles, there is a maternal and a paternal chromosome set but moles arise through a global methylation defect due to mutations in *C6orf221* or *NLRP7* (JUDSON *et al.* 2002; MURDOCH *et al.* 2006; PARRY *et al.* 2011).

1.2.4.3. Histone modifications, chromatin, ncRNAs and enhancer competition in imprinting.

Histone proteins are the major component of chromatin. However, histone modifications have so far been considered secondary epigenetic elements to DNA methylation for the control of imprinted germline DMRs (FERGUSON-SMITH 2011). Indeed, DNA methylation is required for the *establishment* and *maintenance* of imprinting (LI *et al.* 1993). However, some new research suggests that histone modifications could be important players too.

Histones can be modified by methylation, acetylation, ubiquitylation, and phosphorylation. These numerous modifications can alter imprinted gene expression (ZHANG 2003; LEWIS *et al.* 2004; UMLAUF *et al.* 2004; MARGUERON

et al. 2005; MARTIN and ZHANG 2005; THIRIET and HAYES 2005; MARTIN and ZHANG 2007; ROY *et al.* 2010). Furthermore specific histone modifications have been found on the parental alleles of the germinal DMRs (KACEM and FEIL 2009) (see Table 1).

On the methylated ICR allele	**On the unmethylated ICR allele**
HP1 H3K9me3 H4/H2A R3me2s H4K20me3	H3K4me2/3 H3/H4 acetylation

Table 1: Parental-specific histone modification at germinal DMR (KACEM and FEIL 2009).

These histone modifications could create a regulatory code for the control of imprinted DMRs. For example, H3K4 demethylation has been suggested to be important for the docking of DNMT3A and DNMT3L at the germinal DMRs of imprinted genes where these proteins will methylate CpG residues appropriately separated by 6 nucleotides (Figure 4) (CEDAR and BERGMAN 2009; CICCONE and CHEN 2009).

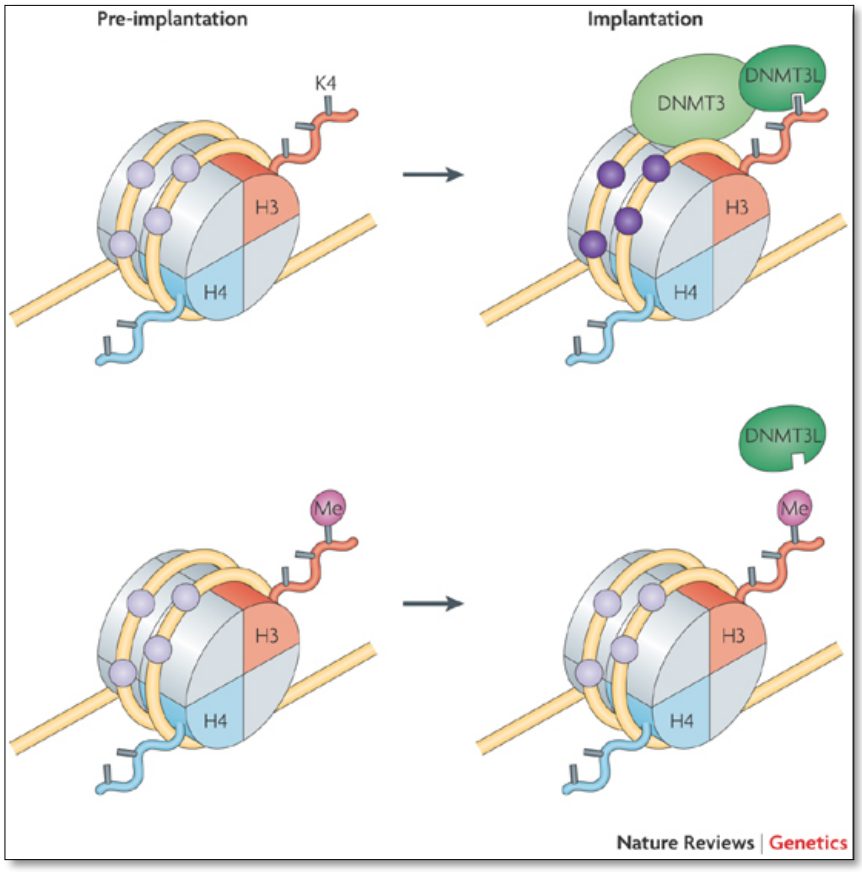

Figure 4: H3K4 demethylation would create docking sites for Dnmt3L/3A tetramer. Figure from Cedar and Bergman (CEDAR and BERGMAN 2009) reused with permission.

The histone modifications present at imprinted gene DMRs could be primary, concomitant or secondary to the DNA methylation of the germline imprinted DMRs. The DMR histone marks can be tissue specific. It has for example been shown that H3K9me2 is specific to the *Kcnq1* imprinted locus in placenta but not in liver (KACEM and FEIL 2009).

Other epigenetic features are important in the establishment and maintenance of imprinting. Polycomb proteins (for example PRC2) seem to be necessary to bring repressive histone modifications to certain imprinted alleles (for a review see (KACEM and FEIL 2009)). It can also be necessary to bring long ncRNAs to regulate the expression of some imprinted genes (ZHAO *et al.* 2010). The regulatory protein ZFP57 has been shown crucial to the correct methylation of certain maternal methylated DMRs as for example *SNRPN*-DMR. Its absence

provokes a variegating imprinted syndrome in mouse and human (LI *et al.* 2008; MACKAY *et al.* 2008).

Insulators have two roles in transcription regulation. They can block the interaction between enhancers and promoters (enhancer-blocking insulators) or create a barrier against the conversion of euchromatin into heterochromatin (barrier insulators) (GILES *et al.* 2010). They have a crucial role at imprinted loci. For example, at the well-characterised *IGF2/H19* locus, the insulator has an enhancer-blocking role on the maternal allele (Figure 5). This blocking is mediated by the binding of CCCTC-binding factor (CTCF) on the unmethylated DMR just 5′ of *H19* promoter (Figure 5) (BELL and FELSENFELD 2000; HARK *et al.* 2000; SZABO *et al.* 2000).

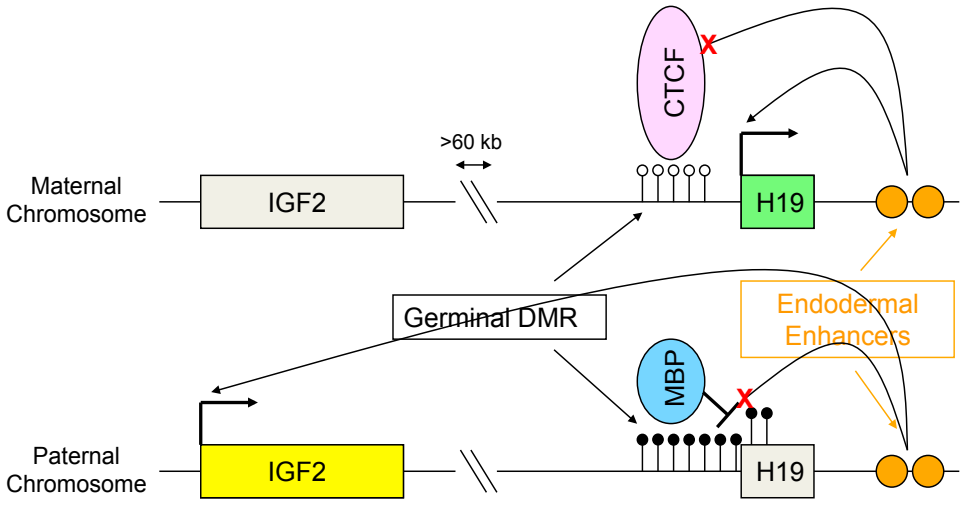

Figure 5: Insulator activity at the *IGF2/H19* locus.
The maternal DMR is not methylated which permits the binding of an insulator protein, CTCF, which will prevent the downstream enhancers from accessing the *IGF2* promoter. The enhancers will then trigger the expression of *H19*. The paternal DMR is methylated. Methylation blocks the binding of CTCF and allows methyl-binding proteins to bind to the paternal DMR and silence the *H19* gene. CTCF being absent, the enhancers can access the *IGF2* promoter, which results in *IGF2* expression.

1.3. Medical relevance of imprinting

1.3.1. **Imprinting syndromes**.

The imprinted syndromes known so far are summarised in Table 2 (http://www.ncbi.nlm.nih.gov/omim).

Table 2: Summary of imprinting syndromes, their phenotypic characteristics, genetic and epigenetic causes and mouse models.

Human imprinting syndromes	OMIM	Clinical presentation	Chr location	Genes	Mechanisms and epimutations	Germline DMRs
Transient Neonatal Diabetes Mellitus	601410	Intrauterine growth restriction, neonatal hyperglycemia requiring insulin therapy, failure to thrive, dehydration, susceptibility to later onset of diabetes	6q24.2	Overexpression of *PLAGL1*, mutations of *ZFP57* (6q22.1), *ABCC8* (11p15.1), *KCNJ11* (11p15.1)	Paternal UPD6 (20%), paternally inherited duplication of 6q24, methylation defect of TNDM DMR (20%)	*HYMAI* exon 1
Silver-Russell syndrome	601523	Severe intrauterine growth restriction, postnatal growth deficiency, body asymmetry, craniofacial abnormalities	complete or partial mUPD7, 7p11-p13, 7q31-qter, 11p15.5	*H19*, *IGF2*, *CDKN1C*, *KCNQ1OT1*	Hypomethylation of ICR1 and ICR2 (50%), maternal UPD 7 (10%)	*H19* DMR (ICR1) and *KCNQ1* DMR (KvDMR1)

Human imprinting syndromes	**OMIM**	**Clinical presentation**	**Chr location**	**Genes**	**Mechanisms and epimutations**	**Germline DMRs**
Beckwith-Wiedemann syndrome	130650	Incidence: 1/14500 live births. Intrauterine overgrowth, placentomegaly, hydramnios, raised AFP, exomphalos, macroglossia, Wilms tumor and other neoplasms, neonatal hypoglycemia, visceromegaly, ear lobe creases, hemihypertrophy, genitourinary anomalies, monozygotic twinning	11p15.5	*CDKN1C* (coding mutation in 5% of sporadic cases and ~40% of dominant cases), *IGF2* (biallelic expression <10%), *H19* (loss of expression biallelic methylation 1-2%), *KCNQ1OT1*(biallelic expression in 50%)	Group I: (20%) UPD, aberrant methylation of *H19* and KCNQOT1 Group II: (7%) aberrant methylation of *H19* Group III: (55%) aberrant methylation at *KCNQ1OT1* Group IV: (18%) No methylation defect detected in *H19* and *KCNQ1OT1* biallelic expression of *IGF2*, loss of methylation at the KvDMR1 (50%), hypermethylation of ICR1 (5-10%) paternal UPD 11p15 <1%	*H19* DMR, KvDMR1
Familial nonchromaffin paraganglioma	168000	Tumors derived from paraganglia, nonchromaffin types primarily serve as chemoreceptors and are located in the head and neck region	11q13	*SDHD*	Mutations of *SDHD*	

Human imprinting syndromes	**OMIM**	**Clinical presentation**	**Chr location**	**Genes**	**Mechanisms and epimutations**	**Germline DMRs**
Maternal and paternal UPD14 syndromes	608149	Paternal UPD14: skeletal abnormalities, joint contractures, dysmorphic facial features, developmental delay, mental retardation.	14	Locus not defined		
Prader-Willi syndrome	176270	Reduced fetal movements, failure to thrive at birth then hyperphagia resulting in early onset obesity, muscular hypotonia, mental retardation, behavioral problems, short stature, hypogonadotopic hypogonadism, small hands and feet	15q11.2-q13	*SNRPN, NDN,* clusters of snoRNAs	De novo deletions of 15q11-q13 on the paternal chr (70%), matUPD 15 (29%) imprinting defect (abnormal methylation) (1%)	PWS-IC (*SNRPN* promoter and exon 1)

Human imprinting syndromes	**OMIM**	**Clinical presentation**	**Chr location**	**Genes**	**Mechanisms and epimutations**	**Germline DMRs**
Angelman syndrome	105830	Incidence 1/15000 live births, mental retardation, movement or balance disorder, characteristic abnormal behaviors, happy disposition, severe limitations in speech and language, hypopigmentation, late-onset obesity	15q11.2-q12	*UBE3A*	De novo maternal deletions (70%), UBE3A mutations (5-10%), patUPD (2%), imprinting defects hypomethylation of SNRPN promoter (3-5%)	PWS-IC (SNRPN promoter and exon 1)
Pseudohypoparathyroidism 1b, Albright hereditary osteodystrophy, McCune-Albright syndrome	603233, 103580	Resistance to PTH: hypocalcemia and hyperphosphatemia, loss of expression of Gs-alpha-protein in renal proximal tubules	20q13.32	*GNAS , STK16*	GNAS mutation or epimutation	*GNASXL* and *NESPAS* promoter DMR (primary ICR), *GNAS* exon 1a promoter DMR (secondary DMR)

1.3.2. Imprinting, medicine and human reproduction.

1.3.2.1. Assisted reproduction technologies (ART) and imprinting.

The use of ART is widespread in humans and farm animals. In 2003, just under 4% of human births in developed countries had been conceived in vitro (ANDERSEN *et al.* 2007). In ART procedures, the oocyte, the zygote and the embryo are subjected to non-physiological conditions, which can potentially lead to epigenetic alterations (and so potentially to dysregulation of imprinted gene expression) (FORTIER *et al.* 2008; MORGAN *et al.* 2008; RIVERA *et al.* 2008; SANTOS *et al.* 2010).

In humans, there have been reports of a very slightly increased incidence of in vitro fertilisation (IVF) or intracytoplasmic sperm injection (ICSI) conceptions first in Angelman syndrome (AS) patients (COX *et al.* 2002; ORSTAVIK *et al.* 2003) and then in Beckwith-Wiedemann syndrome (BWS) patients (DEBAUN *et al.* 2003; GICQUEL *et al.* 2003; MAHER *et al.* 2003). The epigenetic change found in 23 of the 24 cases of BWS after IVF that were analysed and reviewed by Maher was hypomethylation at the *KvDMR1* (MAHER 2005). Interestingly, the change observed in AS cases was also hypomethylation but of the *SNRPN* DMR (COX *et al.* 2002). A study in Denmark has compared the frequency of imprinting diseases in singletons born after IVF with the incidence in naturally conceived singletons (LIDEGAARD *et al.* 2005). In Holland, questionnaires have been sent via support groups to the parents of BWS, PWS and AS patients (DOORNBOS *et al.* 2007). In Ireland and central England, questionnaires have been sent to all couples that had conceived through IVF and some of their children have been invited for a clinical assessment (BOWDIN *et al.* 2007). Overall the findings of these three studies have been reassuring and the absolute risk of imprinting disorders in children conceived through IVF was considered small. There is however a consensus to continue scientific investigations to ensure the safety of new ART protocols (GOSDEN *et al.* 2003; BOWDIN *et al.* 2007).

1.3.2.1.1. *Gametes of infertile parents could be epigenetically disrupted.*

ICSI is an advanced IVF procedure in which a single sperm is injected directly into the oocyte, which is then placed in cell-culture. The use of ICSI can help infertile couples to conceive when the sperm is of poor quality. The methylation level of *H19* was studied in men with normozoospermia, moderate oligozoospermia and severe oligozoospermia (MARQUES *et al.* 2004). The *H19* methylation profile was altered in 0%, 17% and 30% of them respectively (MARQUES *et al.* 2004). Methylation abnormalities might be the cause or the consequence of subfertility.

1.3.2.1.2. *ART could increase LBW through imprinting dysregulation.*

According to Barker's epidemiological work, we know that babies born with a low or very low birth weight (LBW) are at increased risk of short and long term morbidity and mortality (BARKER *et al.* 1989; BARKER *et al.* 2002; BARKER 2006). For this reason it is crucial to better understand the genetic and epigenetic causes of growth restriction. As the maternal age at the time of the first pregnancy increases and as fertility decreases with age (DUNSON *et al.* 2004), the use of ART increases. In singleton pregnancies, the risk of low birth weight (≤2500g) at term following ART was estimated to be 2.6 times that of the general population (95% CI, 2.4 to 2.7) (SCHIEVE *et al.* 2002). The cause of this remains unclear. However, given the relationship of imprinted genes and birth weight, one hypothesis could be that ART provokes mild epigenetic dysregulation that induces low birthweight.

1.3.2.1.3. *Livestock and mouse ART and imprinted-like disease.*

In livestock, a condition called 'large offspring syndrome' (LOS) has been observed in some lambs and calves born after in vitro embryo culture (YOUNG *et al.* 2001). This syndrome is characterised by increased birthweight, organ overgrowth, and neonatal respiratory distress syndrome, reminiscent of BWS in humans. Even though no difference in *IGF2* levels was demonstrated by RT-PCR between large offspring (LO) fetuses and controls (YOUNG *et al.* 2001), the

level of *IGF2R* expression was reduced by 30 to 60% in LOS sheep and was plausibly caused by partial to complete loss of methylation at the normally methylated maternal *IGF2R* DMR.

In mouse, embryo culture in growth media has been shown to affect imprinting (DOHERTY *et al.* 2000; KHOSLA *et al.* 2001). It seems that embryo culture affects mainly the placenta (MANN *et al.* 2004).

1.3.2.2. Placenta, intra-uterine growth and imprinting.

A functional placenta is paramount for the normal growth of the fetus and imprinted genes are crucial for placental development (GEORGIADES *et al.* 2001; REIK and WALTER 2001; SMITH *et al.* 2002; TYCKO 2006). The human placenta is the organ of choice to screen for imprinting as many imprinted genes are expressed in this tissue (COAN *et al.* 2005). Studies on mice have contributed to understand the physiology of this organ and the genetic and epigenetic basis of its (mal)function even though its anatomy differs between mice and humans (Figure 6).

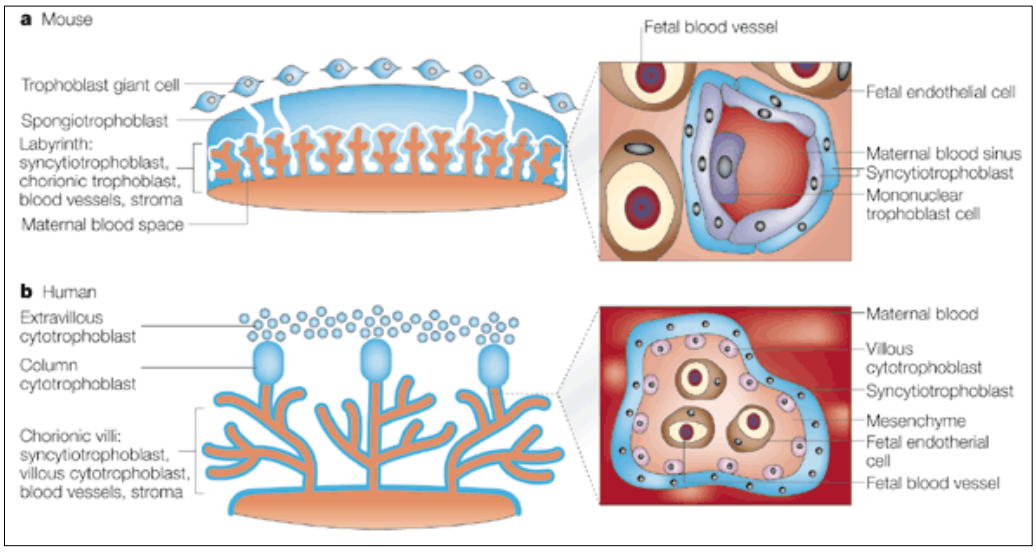

Figure 6: Schematic representation of mouse and human placental development from Rossant and Cross (ROSSANT and CROSS 2001). Reused with permission.

The medical relevance of (placental) imprinted genes might not be restricted to genetic syndromes that represent the extreme end of the ASE deregulation (summarised in section 1.3.1). Mild dysregulation of imprinted genes could create fetal growth disorders (IUGR and macrosomia) (Figure 7). Individual variation due to SNPs or CNVs could also explain the normal distribution of birth weight. One example is given by *PHLDA2*, a maternally expressed imprinted gene on Chromosome 11. Its expression is negatively correlated with birthweight (McMinn *et al.* 2006; Apostolidou *et al.* 2007; Diplas *et al.* 2009; Ishida *et al.* 2012).

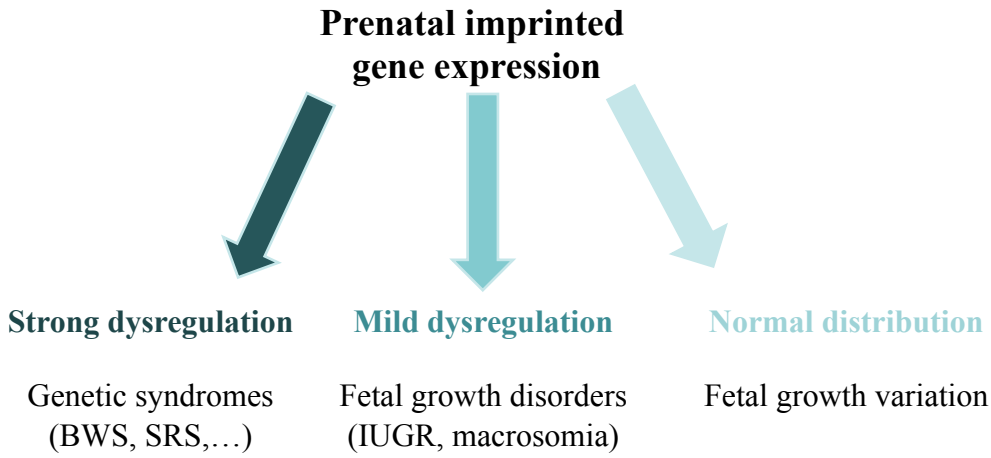

Figure 7: Summary of the phenotype spectrum and medial relevance of dysregulation of imprinted expression.
Individual variation in allelic expression might at best explain some slight phenotypic variations. Strong dysregulations are known to cause some clinical syndromes. They could also influence common diseases, be individual specific and explain some of the missing heritability for common diseases.

1.4. The different types of quantitative variation in transcription and their medical relevance.

1.4.1. Definition of 'biallelic expression'.

By default in diploid organisms, both copies of a gene are functional and produce a RNA transcript. However, some genes (or gene regions) can be present in 0, 1, 2, 3, 4, 5, 6, … copies complicating the basic diploid view of the human genome. These are called copy-number variants (CNV) and occur commonly in 'normal' individuals. Diseases and syndromes can occur when one copy of a gene is missing (deletion), activated or inactivated (mutation), or in a non-tolerated dosage (CNV) which results in an abnormal level of the gene product (typically a protein). There are many diseases and traits that are likely to be caused by CNVs. A few examples are autism (SEBAT *et al.* 2007), HIV susceptibility (GONZALEZ *et al.* 2005), DiGeorge syndrome (ENSENAUER *et al.* 2003), Williams-Beuren syndrome (CUSCO *et al.* 2008), as well as unelucidated cases of Prader-Willi (BUTLER *et al.* 2008) and Angelman (MEFFORD *et al.* 2009) syndromes (see the Online Mendelian Inheritance in Man database for specific syndromes, http://www.ncbi.nlm.nih.gov/Omim).

The development of array-based comparative genomic hybridisation (array-CGH) has enabled a more precise appreciation of the extent of CNV in the normal genome (REDON *et al.* 2006; CARTER 2007). The database of Genomic Variants (http://projects.tcag.ca/variation/) (IAFRATE *et al.* 2004) currently lists 29133 CNVs. This extensive catalogue will help to determine which CNVs are associated with disease in order to implement CGH assays in clinical diagnostic laboratories (LEE *et al.* 2007; LUPSKI 2007; MCCARROLL and ALTSHULER 2007; SCHERER *et al.* 2007). Next generation sequencing has, as expected, permitted an even finer resolution of CNVs (WANG *et al.* 2008a; WHEELER *et al.* 2008; ALKAN *et al.* 2009; CHIANG *et al.* 2009). CNVs are very common and of every size (from few bp insertions-deletions to hundreds of kb). Very rare CNVs are associated with various pathologies, for example obesity (BOCHUKOVA *et al.*

2010), schizophrenia (STEFANSSON *et al.* 2008) or mental retardation (MEFFORD *et al.* 2008).

Clearly a new era of 'biallelic normality' definition is ahead with a 'normal' copy number range to be experimentally determined for each gene. In human, the DECIPHER (http://www.sanger.ac.uk/PostGenomics/decipher/) collaborative project of sharing all information on CNVs of phenotyped patients will enable clinicians to differentiate disease causing CNVs from 'innocuous' ones. Technically, array-CGH, qPCR, sequencing and bioinformatics will be used to fulfil this goal. Clinically, haploinsufficiency or complete absence of gene copies is more often pathogenic than a gain in copy number.

1.4.2. Differential allelic expression of 'biallelic' genes.

Genes can vary in their numbers and such quantitative variation can be medically relevant (see above). Another genetic variability that can also disturb the biallelic norm is the functional non-equivalence of the two alleles. For example some alleles in the population could have SNPs creating a stronger promoter or enhancer, resulting in higher level of expression of the gene transcript and protein.

Variations of the expected 50:50 gene expression between 'biallelic' alleles have been well documented. This appears to affect around 20% of genes and could underlie much of human variability (see Table 3 for a summary of initial studies).

1.4.2.1. Small scale ASE studies.

Yan et al. were the first to report differential allelic expression in human (YAN *et al.* 2002). They studied 13 genes and detected 1.3 to 4.3-fold expression differences between alleles for 6 genes in a minority (3 to 30%) of the 96 lymphoblastoid cell lines (LCLs) from the Centre d'Etude du Polymorphisme

Humain (CEPH) they studied. Bray et al. found ASE in seven out of 15 genes in at least one individual in human brain tissue (BRAY *et al.* 2003). Pastinen et al. studied 129 genes on LCLs and found ASE in 18% of them in more than one individual (PASTINEN *et al.* 2004).

1.4.2.2. Large scale ASE studies.

Lo et al. studied 1494 SNPs in kidney and liver of seven fetuses with an oligonucleotide microarray modified to analyse allele-specific gene expression (Affymetrix HuSNP chip) (LO *et al.* 2003). A subset of SNPs (1063) was located in transcribed regions, 602 SNPs were heterozygous in at least one fetus, 326 (54%) showed preferential expression of one allele, and 170 (28%) showed more than a four-fold difference between the two alleles (LO *et al.* 2003). This report was the first large-scale study of differential allelic expression in human tissues. Subsequently, Pant et al. used oligonucleotide microarrays (Perlegen) to study transcript levels in white blood cells (WBC). They detected allelic expression differences in at least one individual in 53% of the 1389 genes they had targeted (PANT *et al.* 2006).

Such allelic-expression screens are ideal to discover new imprinted genes as these are at the extreme ends of the allelic imbalance spectrum on the human autosomes. For example, *ZNF331* was discovered to be imprinted in WBC by Pant et al. (PANT *et al.* 2006). One limiting factor is the need for commonly expressed SNPs (i.e. with a high heterozygosity in the population studied) to avoid the need to test numerous individuals and so to reduce costs. Another limiting factor is that allelic expression studies have mainly used Epstein-Barr virus (EBV) immortalized LCL of CEPH individuals (PASTINEN and HUDSON 2004; PASTINEN *et al.* 2004; GIMELBRANT *et al.* 2007; POLLARD *et al.* 2008; SERRE *et al.* 2008b) or cultured cancer cell lines (MILANI *et al.* 2007; TAN *et al.* 2008). Cultured cell lines are in general easier to obtain than human tissue for large-scale genetics studies. However, immortalisation using EBV and culture could potentially alter their biology and epigenetic status and in turn, affect the fidelity of the data obtained. To circumvent this issue, several studies have used

native human tissues instead (fetal tissues, primary white blood cells and brain samples) (BRAY *et al.* 2003; LO *et al.* 2003; KNIGHT *et al.* 2004; PANT *et al.* 2006). Both LCLs and native tissue seem to show roughly the same level of differential allelic expression around 20% (see Table 3, see next page). To insure biological relevance, primary tissue (human placenta) has been used in our project.

Table 3: Summary of initial ASE studies.

Study	Material	Technology	ASE	Summary
Yan et al. 2002	LCL of CEPH families 96 individuals	13 genes fluorescence peak/primer extension	6 genes/13 and heritable	23% in more than one individual
Bray et al. 2003	8 to 26 informative subjects (brain samples) (60 samples in total)	15 genes fluorescence peak/primer extension	7 genes/15 in at least one individual	46% of genes in at least one individual
Lo et al. 2003	kidney and liver of seven fetuses	Affy HuSNP 1494 SNPs	326/602	54% in at least one individual 28% of extreme ASE (4 fold)
Pastinen et al. 2004	63 LCLs (Caucasian, Asian, Oceanic and African) + 5 CEPH pedigrees for 9 genes	193 SNPs in 129 genes expressed single base extension with fluorescence	23/126, IL1A, HTR2A FGB random monoallelic	18% in more than one individual
Pant et al. 2006	white blood cells of 12 unrelated individuals	Perlegen custom made 8406 exonic SNPs in 4102 genes	1389 genes studied	53% ASE in at least one individual

Study	Material	Technology	ASE	Synthesis
Pollard et al. 2007	67 CEPH individuals	Perlegen 3877 SNPs in 2625 genes	496 show ASE (two-fold)	13% ASE in at least one individual
Gimelbrant et al. 2007	Clonal B-lymphoblastoid cell-lines, 1mm^3 patches of human placenta	Nsp 250K 11,401 genes 1% in exonic sequence		Random allelic expression in 5% of genes studied
Milani et al. 2007	13 tumor cell lines	Multiplex minisequencing	210 genes, 60 sufficiently expressed	68% ASE in at least one individual
Serre et al. 2008	81 LCLs	Illumina BeadArray	1432 SNPs, 512 above background noise (36%)	17% ASE in at least one individual

1.4.2.3. Large scale ASE-eQTLs studies.

To pinpoint the genetic determinants responsible for ASE (i.e. allelic variation in mRNA abundance), patterns of allelic expression have been studied using linkage strategies (MORLEY *et al.* 2004). Indeed, the mRNA abundance of a gene can be seen as a quantitative trait (called an expression quantitative trait locus-eQTL). Many genetic determinants can regulate these eQTLs such as SNPs, CNVs, haplotypes and epimutations. For example, the regulatory SNPs (rSNPs) influence the expression of nearby genes (cis-acting) or of genes located on another chromosome (trans-acting) and are an important source of phenotypic variability (MORLEY *et al.* 2004; CHEUNG *et al.* 2005; STRANGER *et al.* 2005; DIXON *et al.* 2007; GORING *et al.* 2007; SPIELMAN *et al.* 2007; STRANGER *et al.* 2007; EMILSSON *et al.* 2008).

To find rSNPs, total gene expression is measured in numerous samples using gene-expression arrays and SNPs are genotyped using genotyping arrays. This very simple approach has allowed hundreds of SNPs to be implicated as causal of eQTLs. The same approach has been recently used with RNA-Seq (transcriptome high-throughput sequencing). Again, a lot of eQTLs were found in addition to allelic, isoform specific expression (using HapMap lymphoblasts RNA, parental allelic information was missing) (MONTGOMERY *et al.* 2010; PICKRELL *et al.* 2010; NICA *et al.* 2011).

Functional assays could prove the eQTL 'causality' of a specific SNP but are very expensive and low throughput. Hence, only the addition of several layers of information (epigenomics, evolution conservation, transcriptomics) will elucidate eQTL-rSNP effects further. A collaborative effort is underway to undertake genome-wide expression and methylation studies then transcriptome sequencing in tissue samples of a vast cohort of twins to characterize common trait susceptibility (NICA *et al.* 2011).

1.4.2.4. Epigenetics and ASE.

It is expected that rSNPs provoking ASE will create epigenetic differences at the SNP localisation and/or at the gene they 'control': for example, more or less binding of a TF will bring different levels of cofactor binding and of histone modifications on each allele (Figure 8) (BIRNEY *et al.* 2007; CONSORTIUM *et al.* 2007; HON *et al.* 2009).

It has also been shown that allele-specific epigenetic modifications could be established even if the alleles are genetically similar. The obvious example is imprinted alleles: at these loci, DNA passed through the female germ line will be epigenetically different from the one passed through the male germ line even if the DNA sequences are genetically identical. Finally, random epigenetic noise could create epigenetically different alleles just by 'chance'. Such epigenetic random allelic variation explains some individual, age-dependent or tissue-specific allelic variation even for genetically identical alleles. For example, this type of 'epigenetic drift' has been showed at the chromosomal level in ageing monozygotic twins (FRAGA *et al.* 2005). For both gene copies in the same individual, we could hypothesise that two homozygous alleles might not stay epigenetically identical throughout life. However, *single molecule* sequencing methods will be needed to study epigenetic differences for homozygous alleles.

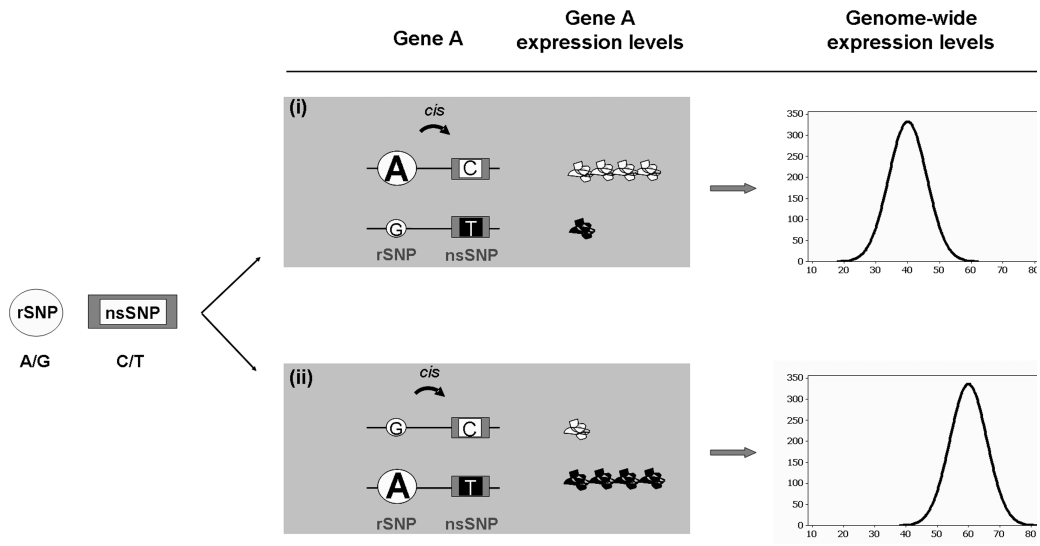

Figure 8: Schematic representation of an interaction between a regulatory and a protein-coding SNP according to Dimas et al. (DIMAS *et al.* 2008).
In two genotypically identical individuals (double heterozygotes), the phasing of alleles can be different and lead to different expression levels of proteins. The A allele will trigger an increased expression of the protein arising from the C allele in cis while the G allele will trigger a low expression from another protein arising from the T allele. This difference can potentially lead to a phenotypic difference (right part of diagram). Figure reused with permission (open access).

In conclusion, ASE of biallelic genes could be a possible cause of biological variation (Figure 8) and may reveal itself medically relevant in the near future. For example, some ASE SNPs were implicated in genome-wide association studies (GWAS) and were functionally proved as rSNPs (KNIGHT *et al.* 2003; KONG *et al.* 2009; NICA *et al.* 2010; SMALL *et al.* 2011).

1.4.3. Random monoallelic expression.

Immunoglobulin genes (PERNIS *et al.* 1965), odorant receptor genes (CHESS *et al.* 1994), T cell receptors genes (RAJEWSKY 1996), interleukins (HOLLANDER *et al.* 1998; RHOADES *et al.* 2000), pheromone receptors (RODRIGUEZ *et al.* 1999) and the protocadherin genes (ESUMI *et al.* 2005) have all been shown to be monoallelically expressed in a random manner (i.e. stochastic expression of either the paternal or the maternal allele).

The random expression of only one allele was first demonstrated in rabbit lymphoid tissues by immunofluorescence (PERNIS *et al.* 1965). In mouse odorant receptor genes, Chess et al. have shown by different methods that odorant receptors are monoallelically expressed. By RT-PCR of cell pools in a specific neuron, they showed that the I7 receptor gene is expressed from one allele only (CHESS *et al.* 1994). In S phase, both alleles of the vast majority of genes are replicated synchronously (HOLMQUIST 1987). While for genes monoallelically expressed (IG, X chromosome genes), alleles are asynchronously replicated (TAYLOR 1960; KITSBERG *et al.* 1993; KNOLL *et al.* 1994).

More recently, Gimelbrant et al. surveyed clonal human B-lymphoblastoid lines for random monoallelic expression using the 250K SNP array (Affymetrix) with a protocol modified to enable cDNA genotyping. SNPs were analysed in 3939 genes in two or more clonal cell lines. A total of 371 genes (9.5%) were dubbed 'randomly monoallelically expressed' in most clonal cell lines but not all. For more than four-fifths of monoallelic genes, some clonal cells exhibited biallelic expression (GIMELBRANT *et al.* 2007). Many of the monoallelically expressed genes were highly expressed in the cell lines studied which was reassuring regarding the potential stochastic monoallelic expression observed at low levels of transcription (OHLSSON *et al.* 1998).

One hypothesis to explain this phenomenon could be that for some genes, during the development of an organ, a random epigenetic allelic bias is created and so a cell (and all its descendants) starts to express one allele more favourably than the

other. As for random X-inactivation, only strong skewing of expression towards one allele *in the whole organ* would create a medically relevant allelic bias. So while random monoallelic expression is conceptually important, its random nature makes it probably prone to 'natural buffering' and a priori a more rarely relevant medical actor. We can foresee an exception to this statement: if one of the alleles is null (or weaker), an organ could end up with some cells expressing the gene and some not (or less). In this case, some variation inside an organ could exist and could be biologically relevant.

How is random monoallelic expression established? The mechanisms of allelic exclusion in immunoglobulins closely resemble random X-inactivation. It was found to be established before B-cell development (MOSTOSLAVSKY *et al.* 2001). These sets of experiments suggest that epigenetic marks control the allelic exclusion process. Integrative studies (RNA-Seq, ChIP-Seq) using ultra sensitive single molecule sequencing methods (starting from a few cells) will be needed to characterise in depth the epigenetic differences causing monoallelic expression.

The allele-specific modes of expression described above were all limited to the expression of one gene. However, a more complex mode of ASE due to switching off a complete chromosome or locus is called heterochromatinisation.

1.4.4. Chromosomal or whole locus heterochromatinisation.

1.4.4.1. X-chromosome inactivation.

X-chromosome inactivation (XCI) is the random transcriptional silencing of most genes on one of the two X chromosomes in female cells. Ohno and Hauschka showed that one X-chromosome was heteropyknotic in normal diploid cells in female mice (OHNO and HAUSCHKA 1960). Mary Lyon later suggested that the heteropyknotic X-chromosome was inactivated and so in a random fashion (LYON 1961). The randomness of XCI was shown by the tortoiseshell coat colour in female mice heterozygous for the normal and mutant colours and

by the mosaic phenotype in female mice heterozygous for other X-linked mutants (LYON 1961).

In placental mammals, X inactivation is usually random (i.e. either the paternal or the maternal chromosome is inactivated in different cells) whereas in marsupials it is always the paternal copy which is inactivated (CHOW and HEARD 2009). In mouse, the paternal X-chromosome is inactivated in extra-embryonic tissues (imprinted X-inactivation) (TAKAGI and SASAKI 1975). In the inner cell mass of the blastocyst, the imprinted X-inactivation is 'reset' and in each cell, the inactivation becomes random. It is the X-inactivation center (XIC) that controls the X-inactivation process. It encodes two non-translated RNA genes: the long X-inactive specific transcript (XIST) (BORSANI *et al.* 1991); (BROCKDORFF *et al.* 1991; BROWN *et al.* 1991) and its antisense *Tsix* (LEE *et al.* 1999a). In mouse, *Xist* has been shown to 'coat' the future inactive X chromosome (Xi), from which it is expressed, and impede its transcription (HERZING *et al.* 1997). *Tsix* is responsible for the silencing of *Xist* on the future active X-chromosome through chromatin structure modification (NAVARRO *et al.* 2005; SADO *et al.* 2005; SUN *et al.* 2006). The Xi will exhibit accumulation of transcripts from the *XIST* gene (CLEMSON *et al.* 1996) and of repressive histone modifications: deposition of H3 lysine-27 trimethylation (H3K27me3) and depletion of H3K4me3 (MARKS *et al.* 2009). To try to elucidate the role of methylation in X inactivation, a recent study has compared the methylation of X chromosomes in normal females (46,XX) and Turner syndrome patiens (45, X) using methylated DNA precipitation (MeDIP) and subsequent hybridisation to a high-density oligonucleotide array (SHARP *et al.* 2011). Sharp et al. note that XCI correlates with gain of methylation at the majority of CGIs of Xi genes and reduced methylation that occurs at a number of CGIs correlates with XCI escape (SHARP *et al.* 2011).

1.4.4.2. X-inactivation escape in human.

In human, approximately 15% of X-linked genes escape silencing to some degree on Xi. A further 10% of genes have variable patterns of inactivation and

are expressed at different levels from inactive X chromosomes (CARREL and WILLARD 2005). Carrel and Willard tested 94 genes using a quantitative assay with fluorescent single-nucleotide primer-extension to compare allelic ratios in cDNA and gDNA on 40 human fibroblast samples. The Xi genes escaping inactivation localise to specific chromosomal regions predominantly on the distal portion of X chromosome short arm (Xp) called the pseudoautosomal region. The distribution of pseudoautosomal genes suggests that aneuploidy for Xp is more severe than aneuploidy for Xq. They also noticed that even for pseudoautosomal genes, expression on Xi compared to Xa was variable between individuals (CARREL and WILLARD 2005). Surprisingly, imprinted genes on the X chromosome have also been described in female mice whole brain (DAVIES *et al.* 2005; RAEFSKI and O'NEILL 2005; KOBAYASHI *et al.* 2006).

1.4.4.3. Medical relevance of ASE of X-chromosome genes.

One can expect to find medically relevant X-linked ASE effects in males. Males have only one X chromosome. When the allele present on the X is 'deficient', X-linked disease can arise. One could then postulate that variation in expression levels of X alleles should bring phenotypic variation in males.

Females are mosaics of two cell populations: in one cell population the paternally inherited X chromosome is active and the maternal one is in the other. This will greatly limit the ASE effect *at the organ level*. Indeed, in an organ, the relative proportion of the two cell populations should be close to 50/50 so both alleles will buffer each other. Sometimes the X-inactivation process is skewed (i.e. differs from 50:50) and generates, in females, phenotypic variation or even pathologies only seen in males (ORSTAVIK 2009). In general, female carriers of X-linked conditions are only mildly affected if at all (see Orstavik for review (ORSTAVIK 2009)). Skewing is more prevalent as females become older (8% of extreme skewing at 73 years of age) suggesting epigenetic deregulation with ageing (ORSTAVIK 2009).

1.4.4.4. Other forms of heterochromatinisation.

X chromosome inactivation is the most widespread allelic mode of expression as it extends to a whole chromosome. On a locus level, imprinted genes aside, we need to mention ribosomal DNA (rDNA) loci. Ribosomal DNA loci are scattered throughout the genome and only half of the gene copies are transcribed (GRUMMT 2007). The other half are silenced by noncoding RNA (MAYER *et al.* 2006; SANTORO *et al.* 2010) and by DNA methylation within the promoter region (SANTORO and GRUMMT 2001). These loci show later replication timing (LI *et al.* 2005) and the rDNA genes are repressed in a stochastic manner similar to the X chromosome inactivation process (SCHLESINGER *et al.* 2009). Diseases related to skewed inactivation of ribosomal genes have not yet been described to my knowledge.

1.5. Current count of imprinted genes and screening methods.

1.5.1. Current count.

Catalogues of human and mouse imprinted genes (IG) are regularly updated and are accessible at http://igc.otago.ac.nz/home.html (MORISON and REEVE 1998; MORISON *et al.* 2005), at http://www.mgu.har.mrc.ac.uk/, at https://atlas.genetics.kcl.ac.uk and at http://www.geneimprint.com.
The last formal census of IG counted 71 in mouse, 41 in human and 29 common to both species (http://igc.otago.ac.nz/home.html) (MORISON *et al.* 2005). This number is approximate and changes regularly as new IGs are found or as the imprinting status of other genes is refuted.

For 23 of the 71 imprinted mouse genes, there were no reports of imprinting studies in human (MORISON *et al.* 2005). Twelve genes are imprinted in mouse but have been shown to be biallelically expressed in human (MONK *et al.* 2006; SCHULZ *et al.* 2006; MONK *et al.* 2008). The only gene that is imprinted in human and not in mouse is *L3MBTL*, a polycomb family member located on chromosome 20 (LI *et al.* 2004). Several genes are specific to mouse or human. Selective pressures are different on both species and might account for the differences observed.

The total number of imprinted genes, in either mouse or human, is still not fully ascertained. New imprinted genes continue to be found. An initial estimate of ~100 imprinted genes in the murine genome was derived from restriction landmark genome scanning (RLGS) with methylation sensitive enzymes (HAYASHIZAKI *et al.* 1994). 0.2% of CpG islands showed parent-of-origin specific methylation in this study.

Barlow made another conservative estimate of 100 to 200 imprinted mammalian genes based on the proportion of mouse loci exhibiting parental effects (BARLOW 1995; WATANABE and BARLOW 1996). Morison's opinion is that we already

know most of the human imprinted genes, based on the fact that the causative gene is known for most phenotypes with imprinting features or with a mode of inheritance compatible with imprinting (MORISON *et al.* 2005). However, recent bioinformatics predictions were much less conservative (see below).

Smith et al. have reviewed the different methods used to identify imprinted genes up to the start of this project (SMITH *et al.* 2004). A recent exhaustive review of screens has been performed by Henckel and Arnaud (HENCKEL and ARNAUD 2010). These screens are outlined below.

1.5.2. Historic 'functional' screens for imprinted genes.

1.5.2.1. Direct subtraction screens.

Engineered embryos with chromosomes of entirely maternal (gynogenetic) or paternal (androgenetic) origin have been used to obtain mRNA (DEAN 2001). The subtractive-hybridisation method of gynogenetic and normal embryo cDNAs has led to the discovery of several paternally expressed genes: *Peg1/Mest*, *Peg3/Pw1*, *Peg5/Nnat*, *Peg8/Igf2as* (KANEKO-ISHINO *et al.* 1995; KUROIWA *et al.* 1996; KAGITANI *et al.* 1997; KANEKO-ISHINO 1997). Similarly, the subtraction of androgenetic and normal embryos highlighted: *Meg1/Grb10* and *Meg3/Gtl2* as being maternally expressed (MIYOSHI *et al.* 1998; MIYOSHI *et al.* 2000).

1.5.2.2. cDNA microarrays screens.

Several new IG have been discovered using mice with uniparental duplications or disomies (CHOI *et al.* 2005; SCHULZ *et al.* 2006). The use of microarrays on specific tissues has enabled Schulz et al. to test the whole of mChr 7 (containing the BWS orthologous region) and 11 (containing PWS/AS orthologous region). Four brain-specific paternally expressed transcripts and three placenta-specific

maternally expressed genes were validated by RT-PCR and SNPs (SCHULZ *et al.* 2006).

Mizuno et al. discovered the new IG: *Asb4*. *Ata3*, and *Dcn* using the RIKEN 19K cDNA microarray (18,816 sequences) on parthenogenetic and androgenetic mouse embryos (MIZUNO *et al.* 2002). They tested the candidate genes that had a 'highly differential' expression i.e. the cut-off was the value obtained for known IG on the array (MIZUNO *et al.* 2002).

One other study extended the method and suggested a greatly increased number of candidate IG (above 2000) using expression profiling of full length mouse cDNAs from the FANTOM2 set comprising 18,609 transcriptional units (NIKAIDO *et al.* 2003). These authors discovered two non-coding RNAs mapped to the PWS locus but did not explore the other candidates (NIKAIDO *et al.* 2003). As with the subtraction hybridisation screens, these array studies have been criticised because their expression profiling was conducted on tissues of androgenetic and parthenogenetic mouse embryos. Hence, maternally or paternally expressed IG are respectively missing and are probably deregulating many other genes (MIZUNO *et al.* 2002). Moreover, their method failed to highlight known IG (MORISON *et al.* 2005).

In an elegant study using pyrosequencing to test 64 candidate genes from the Nikaido et al. study (NIKAIDO *et al.* 2003), Ruf et al. questioned the efficiency of expression profiling studies of uniparental mouse embryos (RUF *et al.* 2006). They limited their study to genes located in known imprinted regions in mouse (http://fantom2.gsc.riken.jp/EICODB). Three transcripts were shown to most likely belong to the transcription units of already known IG and the other ones were biallelically expressed (RUF *et al.* 2006).

1.5.2.3. Methylation screens.

Various screens have used the differential methylation on the parental alleles to identify new IG. In mouse, Hayashizaki et al. have used methylation-sensitive

restriction endonucleases (RLGS) to screen the mouse genome for differential methylation at CpG sites and discovered the imprinted U2af binding protein on mChr 11 (HAYASHIZAKI *et al.* 1994). Genes within the complex *Gnas* locus (*Gnas*, *Nespas*) were discovered using methylation-sensitive representational difference analysis on engineered mouse embryos with uniparental disomies (KELSEY *et al.* 1999; PETERS *et al.* 1999). More recently, a screen for maternal methylation (using methylation-sensitive representational difference analysis on parthenogenetic mouse embryos) permitted the discovery of *Peg13*, *Nap1l5* and *Slc38a4* (SMITH *et al.* 2003).

In human, *GNAS* (XLalphaS) (HAYWARD *et al.* 1998) and *PLAGL1* (also called *ZAC* and *LOT1*) (KAMIYA *et al.* 2000) have been isolated also using RLGS on human gynogenetic chimaera (STRAIN *et al.* 1995) and complete hydatiform moles (androgenetic material). Strichman-Almashanu et al. have screened the human genome for normally methylated CpG regions using a restriction enzyme-based strategy (STRICHMAN-ALMASHANU *et al.* 2002). This approach has enabled the discovery of one novel imprinted gene on hChr 18q21: *TCEB3C*, a transcription elongation factor. *TCEB3C* was confirmed imprinted and maternally expressed by RT-PCR on the tissues of four informative fetuses (lung, brain, placenta, and spinal cord) (STRICHMAN-ALMASHANU *et al.* 2002).

1.5.3. Genome-wide bioinformatics predictions of new imprinted genes and functional screens.

Greally et al. had observed that imprinted genes were deficient in short interspersed transposable elements (SINEs) and had asked whether this had been selected for because SINEs are prone to methylation (GREALLY 2002). The content of SINEs in the flanking regions of both monoallelic and imprinted genes was lower than the genome average (GREALLY 2002; KE *et al.* 2002; ALLEN *et al.* 2003).

The density of long interspersed nuclear elements (LINE)-1 transposon sequence have been implicated in X-inactivation. Allen et al. examined whether LINEs were also densely flanking monoallelic or imprinted genes. Their statistical analysis of repeat content revealed a significant higher density of LINEs around random monoallelic genes compared to biallelically expressed genes. However, it was not the case for most imprinted genes (ALLEN *et al.* 2003). Wood et al. have screened the mouse genome for imprinted candidate genes using three criteria: being retrotransposed from Chr X, located within an intron of another gene and overlapped by a CpG island (WOOD *et al.* 2007). They confirmed one of their candidates (Mcts2) to be imprinted (WOOD *et al.* 2007).

Statistics on imprinted locus DNA sequence features (such as CpG islands, repetitive elements and transcription factor binding sites) were performed by Jirtle's group (LUEDI *et al.* 2005). They observed that the relative orientation of repeated elements in the flanking regions of a gene could be important in predicting whether a gene is imprinted or not. They have trained a statistical model to try to predict imprinted genes and the parent-of-origin of the expressed allele firstly in the mouse genome. On the 23,788 annotated autosomal mouse genes (http://www.ensembl.org; version 16.3), their algorithm predicted 600 putative imprinted genes (2.5%) and 64% of them were predicted to be maternally expressed. They found that the orientation of SINEs was of significant weight in their algorithm (LUEDI *et al.* 2005). The bulk of the candidates tested in this project were the human orthologous genes of these mouse imprinted candidate genes (see Tables 1 and 2 'mouse candidates' in Appendices Chapter).

Several groups have used publicly available databases of expressed sequence tags (EST) to predict candidate imprinted genes (YANG *et al.* 2003; GE *et al.* 2005; LIN *et al.* 2005). Using dbEST (version 146), Seoighe et al. have used statistical models in a maximum likelihood framework to detect unequal representation of alleles in EST sequences in cDNA libraries (SEOIGHE *et al.* 2006). Data was sufficient to test about 1,900 genes. Five genes in their set were known imprinted genes. *IGF2* was not found to be statistically significant but

there was a significant enrichment in imprinted genes (two-sided Fisher's exact test; p=1x10-7). Seventy-six SNPs (from 60 different genes) fitted their imprinting model better than the null hypothesis out of a total of 3,969 SNPs for which there were at least five ESTs derived from the less frequent allele. These candidates were incorporated in our list of genes to test (see Tables 1 and 2 'human candidates' in Appendices Chapter).

The total number of imprinted genes is still an open question and thanks to a cohort of placental tissues with matched parental blood samples, we have the necessary resources to test the predictions made by Luedi et al. (LUEDI *et al.* 2005) and Seoighe et al. (SEOIGHE *et al.* 2006).

1.6. Aims of this thesis.

The aim of this thesis was two fold: first the discovery of new imprinted genes in the human placenta; and second the exploration of the epigenetic control mechanisms of previously unstudied or little studied imprinted loci found through these screens.

This thesis discusses:

1) Two medium-throughput technologies (Sequenom® and Illumina®) that were used to test hundreds of candidate genes that were predicted by the bioinformatics studies of Luedi et al. and Seoighe et al. (Chapters 3 and 4).

2) The variability of the silencing of the non-expressed allele for human imprinted genes and the extent of allelic differential expression in the term human placenta (Chapter 3 and 4).

3) The partial imprinted expression of the *PHACTR2* gene in the *PLAGL1/ZAC* locus on chromosome 6 (Chapter 5).

4) The new *ZNF331/C19MC* primate specific imprinted locus on chromosome 19 and its differentially methylated regions (Chapter 5).

Chapter 2. Materials and Methods.

2.1. Samples.

2.1.1. Tissue and DNA resources.

2.1.1.1. Placental tissue samples and matched parental blood samples.

Human placental DNA and RNA and parental genomic DNA trio samples consisting of multiple site placental biopsies with corresponding maternal and paternal blood EDTA samples were previously collected at Hammersmith, Queen Charlotte's and Chelsea Hospital (Apostolidou et al, 2007). Ethical approval was obtained from Hammersmith, Queen Charlotte's and Chelsea and Acton Hospitals Research Ethics Committee, Project registration No 2001/6029. Biopsies were taken on the fetal side of the placenta around the base of the umbilical cord. All samples were washed in sterile PBS, snap frozen in liquid nitrogen and stored at -80ºC until further use. The parents were all of European ancestry. Clinical information such as birth weight, head circumference and gestational age were collected from all newborns in addition to maternal medical details. Both neonatal and maternal records were anonymised while kept linked to each other. The set of trios used in this work were chosen at random in the tissue bank.

2.1.1.2. First trimester trophoblast and matching maternal blood.

A cohort of placentas from first trimester terminations of pregnancy with the matching maternal DNA was also previously collected at Hammersmith, Queen Charlotte's and Chelsea Hospital and Acton Hospitals Research Ethics Committee, Project registration No 2001/6028. A set of 12 trophoblast samples from gestations ranging from eight to 14 weeks was chosen at random from the tissue bank.

2.2. DNA and RNA preparation.

2.2.1. DNA extraction from peripheral blood.

Parental DNAs were previously extracted from peripheral blood using standard phenol/chloroform extraction. For this study, DNA normalised stock solutions (100 ng/μl) and working dilutions (20 ng/μl) were made and stored at 4ºC until use.

2.2.2. DNA extraction from placental tissue.

Pieces of frozen placental tissue (approximately 100 mg) were homogenised with a tissue homogeniser (Ultra Turrax, IKA laboratories, Germany) in 400 ml of 1XTris-EDTA (TE). The homogenised tissues were treated with proteinase K (5 μl at 20 mg/ml) in 5 μl of 10% SDS to lyse the cells and the nuclei and detach the DNA from the chromatin. One hundred μl of 0.5 M EDTA were added before thorough mixing. After incubation overnight at 52°C, DNA was extracted with phenol to remove proteins. Centrifugation at 10,000 rpm for 5 minutes at 16°C separated the aqueous (DNA) and phenol (protein) phases. The upper aqueous layer was carefully transferred into a fresh tube. The procedure was repeated. The remaining phenol was then removed from the aqueous phase by addition of an equal volume of chloroform. The aqueous and chloroform phases were again separated by centrifugation. The aqueous phase was recovered and the procedure repeated. The upper aqueous layer containing highly purified DNA was recovered and DNA was precipitated with 1ml of 100% ethanol and spinning at 10,000 rpm for 5 minutes. The DNA pellet was washed with 500 μl of 70% ethanol and air-dried before being resuspended overnight in TE (100 μl).

2.2.3. RNA extraction from 1st and 3rd trimester placental tissues.

RNA was extracted directly from approximately 100 mg of snap frozen tissue samples using Trizol® (Invitrogen™). Placental tissue samples were homogenised with a tissue homogeniser (Ultra Turrax, IKA laboratories, Germany) in 1 ml Trizol. The Trizol/sample solution was transferred to 1.5 ml eppendorf tubes and 200 µl of chloroform were added. The tube was vortexed thoroughly. The sample was incubated at room temperature for five minutes and centrifugated at 13,000 g for 15 minutes at 4ºC. Four hundred µl of the upper aqueous phase were precipitated in 0.8 volumes of isopropanolol, incubated for 10 minutes at room temperature and centrifuged again at 13,000 g for 15 minutes at 4ºC.

2.3. RNA purification.

2.3.1. DNAse treatment.

Ten µg of total RNA were treated with the TurboDNA-free® kit (Ambion) in a total reaction volume of 50 µl, for 30 minutes at 37°C to cleave the contaminating DNA in small fragments. The DNase was inactivated with 0.1 volume of the DNase inactivation reagent (Ambion) by incubation at room temperature for two minutes with occasional mixing. The tube was then spun at 10,000g for 15 minutes at room temperature. The supernatant was transferred into a fresh tube and stored at -80ºC until use.

2.3.2. Further RNA purification.

The RNA samples treated with the TurboDNA-free® kit (Ambion) were further cleaned from proteins and DNA and concentrated with MinElute™ columns

(Qiagen) according to the manufacturer's instructions. The only modification was the elution in a higher volume of 26 µl of DEPC water (Invitrogen).

2.4. RNA and DNA quantification.

2.4.1. Spectrophotometry.

Extracted and purified nucleic acid samples were quantified by spectrophotometry using a NanoDrop ND-1000 (Nanodrop Technologies, Thermo Scientific). One to 1.5 µl of sample was pipetted onto the pedestal. The spectral measurement (from 220 nm to 75 nm) allowed the automatic quantification of the sample with RNA and DNA correction factors. The pedestal was cleaned with a dry laboratory wipe before a droplet of the next sample was loaded.

2.4.2. Fluorescent nucleic acid stain.

Quant-iT™ RiboGreen® (Invitrogen) RNA quantitation and Quant-iT™ PicoGreen® (Invitrogen) dsDNA quantitation reagents were used to determine the nucleic acid concentrations of the samples subsequently assayed on the Illumina ASE array. A 200-fold dilution of the Picogreen reagent (working solution) was first prepared with 1X TE. A dilution series was then prepared in triplicate with the experimental RNA (ribosomal RNA standard) or DNA stock solution (lambda DNA standard), both diluted 50-fold in TE to make the 2 µg/ml working solution, and TE, in a black 96-well Greiner bio-one plate, as follows:

Volume (µl) of TE	Volume (µl) of RNA or DNA stock solution (2 µg/ml)	Volume (µl) 200-fold diluted RiboGreen or PicoGreen reagent	Final RNA or DNA amount in assay (ng)
0	100	100	200
50	50	100	100
90	10	100	20
98	2	100	4
100	0	100	0

The fluorescence of the standard curve plate was measured in a standard spectrofluorometer DTX 800/880 Series Multimode Detector (Beckman Coulter) thanks to fluorescein excitation and emission wavelengths. Nucleic acid samples (prepared as specified hereunder) were then quantified in duplicate in a black 96-well Greiner bio-one plate. At least two wells were left blank as a negative control.

1) For RNA samples:

Component	Volume
RNA	2 µl
TE	98 µl
RiboGreen working solution	100 µl

2) For DNA samples:

Component	Volume
DNA	1 µl
TE	99 µl
PicoGreen working solution	100 µl

The plates were sealed, wrapped in foil to protect the reagent from light, vortexed gently, spun down and incubated for 2 to 5 minutes at room temperature. The fluorescence was then measured using the same plate reader DTX 800/880 Series Multimode Detector (Beckman Coulter) and the same parameters as for the standard curve. The concentration of samples was obtained by averaging the duplicate fluorescence readings and linear regression analysis using the standard curve.

2.5. Complementary DNA synthesis.

2.5.1. First-strand cDNA synthesis.

The cDNA was synthesised from 250 ng of RNA using Superscript III® reverse transcriptase (RT) (Invitrogen) and random hexamers in a 20 µl reaction for the experiments using the Sequenom platform.

First strand cDNA synthesis reaction as follows:

Component	Amount
Total RNA	250 ng
Random hexamers (3 µg/µl)	0.1 µl
dNTP mix 10 mM	1 µl
H_2O	to 13 µl

The mixture was heated in a microcentrifuge tube to 65ºC for 5 minutes and then promptly incubated on ice for at least 1 minute. The content of the tube was collected by brief centrifugation and the reverse transcription carried out with:

Component	Volume
5x First-strand buffer	4 µl
DTT (0.1 M)	1 µl
RNaseOUT™	1 µl
SuperScript™ III RT	1 µl

The content of the tube was gently mixed (by pipetting up and down) before being incubated at 25ºC for 5 minutes and then at 50ºC for 1 hour. Heating at 70ºC for 15 minutes inactivated the reaction. Samples with the same amount of total RNA but without RT (called RT- samples subsequently) were prepared in parallel with this protocol.

2.5.2. Genomic contamination minimization.

All RT- samples have been run to monitor genomic contamination and the successive steps described earlier (see section 2.3) and detailed in Figure 9, have proven necessary to bring contamination below the detection threshold on the Sequenom platform.

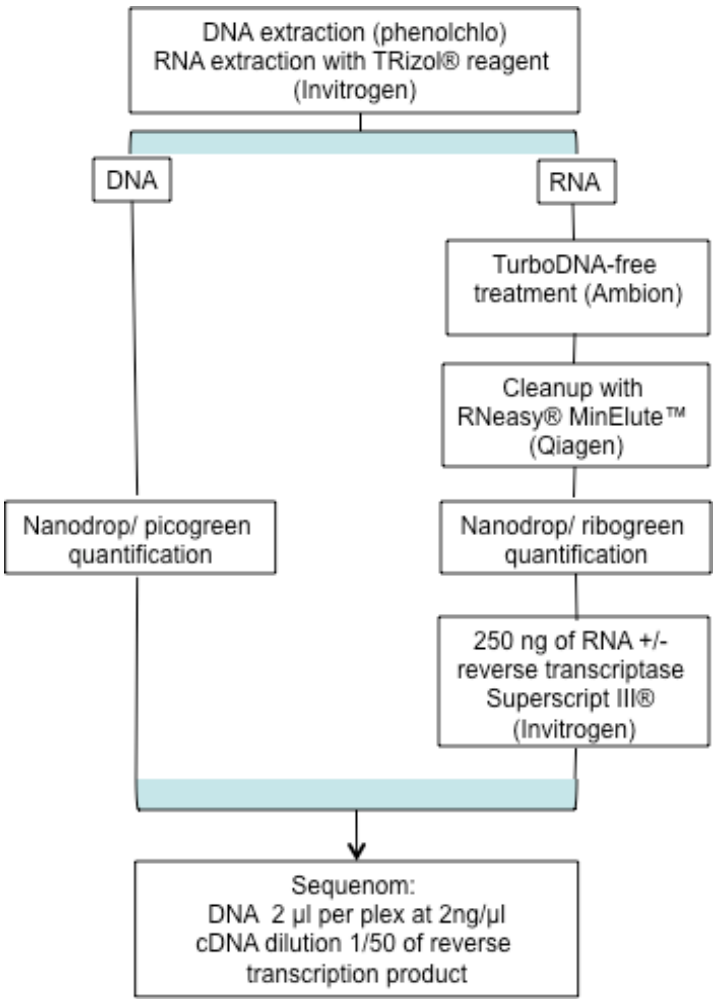

Figure 9: Final protocol for gDNA and RNA extraction, cleaning and quantification.

Complementary DNA samples and their RT- counterparts have been diluted 1/50 before being assayed on the Sequenom platform. All of RT- samples have been run for a subset of the candidates and they all appeared free of genomic contamination once this cleaning protocol was set-up. The samples with RT- controls that appeared free of genomic contamination were subsequently run on the Illumina platform.

2.5.3. Double-stranded cDNA synthesis.

Double stranded cDNA was synthesised for the Illumina experiments according to the manufacturers instructions without taking the DNA strand orientation into account (to increase primer design success). Fresh aliquots of total RNA (250 ng) have been dried down and random hexamers were used to synthetize the first strand. First, the following were combined on ice in an ABgene 96-well plate:

Component	Volume
RNA	250 ng
Random hexamers	0.5 µl
Buffer annealing	0.5 µl
H_2O	3 µl

After an incubation of 10 minutes at 70ºC and one minute on ice, the content of the plate was collected by brief centrifugation. Second, the reverse transcriptase was added:

Component	Volume
2x First strand reaction mix	5 µl
Superscript III RT	1 µl

The samples were vortexed gently and collected by centrifugation before being incubated at room temperature for 10 minutes. The plate was then placed in a thermal cycler for the next two steps: 50ºC for 5 hours and 85ºC for 5 minutes. Once the programme was finished, the plate was placed on ice to terminate the

reaction before the synthesis of the second strand with DNA polymerase I (Invitrogen) and ribonuclease H (Invitrogen).

Component	Volume
$MgCl_2$ (17.5 mM)	4.0 µl
dNTPs mix (25 mM)	0.4 µl
DNA polymerase I	0.6 µl
RNase H	0.2 µl
H_2O	4.8 µl

The reaction was incubated at 16ºC for 5 hours then at 75ºC for 10 minutes before being terminated at 10ºC. The plate was stored at -80ºC until clean-up step.

2.5.4. Cleaning of cDNA samples for the Illumina assay.

This additional cleaning step of cDNA samples has been used for the plates run on the Illumina platform. The cDNA plates were cleaned using a Multiscreen® $PCR_{\mu 96}$ filtration plates (Millipore) to remove primers and unincorporated dNTPs. Water was added to the samples to reach a final volume of 100 µl. The samples were transferred into the Mutiscreen plate and spun at 4000 rpm for 10 minutes. The flow through was discarded. Twenty µl of water were added to each well and the plate shaken for ten minutes at 1100 rpm. The supernatant was transferred into a new 96-well plate. The cDNA samples were dried down before being re-suspended in 5 µl of water.

2.6. Bioinformatics methods.

2.6.1. Choice of SNPs.

Imprinted or monoallelic expression can be detected by exploitation of polymorphisms in the spliced transcript, for example SNPs occurring within the exon or UTR of any gene. These SNPs are used to quantify the expression of each allele in heterozygous individuals. If the placental genomic DNA is heterozygous for a SNP within the gene under study and if at least one parent is homozygous for this SNP, then the parental origin of the alleles and any type of differential allelic expression, imprinted, allelically biased, or random monoallelic can be detected.

Theoretically, a heterozygous genotype for the placenta combined to a homozygous genotype in at least one parent (= an informative family) should be retrieved for 22% of the placenta-parents trios for SNPs with a minor allele frequency (MAF) above the 15% cut-off. Moreover, five or more informative trios are needed to discriminate imprinted expression status (i.e. all samples exhibit either maternal or paternal expression) from random monoallelic expression (i.e. different samples show different parental origin of the expressed allele), with a *p*-value <0.05. Therefore, five divided by 0.22 being equal to 22.72, at least 23 trios should be tested to have sufficient power to detect new imprinted genes.

To be able to study differential allelic expression, SNPs were chosen that were located within the exons or within the untranslated regions (UTRs) of the gene studied. To have as many informative families in our population of European ancestry as possible, SNPs with a MAF above 15 % were selected. For each gene, the SNPs with the biggest MAF were manually picked using dbSNP (Builds 125 and 126) and UCSC databases (http://www.ncbi.nlm.nih.gov/sites/entrez?db=Snp, http://genome.ucsc.edu/).

Moreover, to be studied further, the candidate genes had to be expressed in the human placenta. Therefore their reported expression status was verified in the Unigene database (http://www.ncbi.nlm.nih.gov/entrez/).

2.6.2. Choice of candidate genes expressed in the human placenta.

2.6.2.1. Human imprinted genes and other control genes.

Genes with a known biallelic or imprinted expression pattern were chosen as controls (Table 4, see next page) .

2.6.2.2. Orthologues of mouse imprinted genes

The first candidates to be chosen were the human orthologues of mouse imprinted genes whose status was unknown in human and genes with provisional imprinting data (Table 5, see page 84 and 85).

Table 4: List of control genes

SNP	Gene	Expression pattern in human	Chr loc.	Description
rs13073	*PEG10*	Imprinted (ONO *et al.* 2001)	7q21.3	Paternally expressed gene 10
rs10509305	*STOX1*	Biallelic (IGLESIAS-PLATAS *et al.* 2007)	10q22	Storkhead box 1
rs1063224	*INPP5F*	Biallelic (WOOD *et al.* 2007)	10q26.11	Inositol polyphosphate-5-phosphatase F
rs1057769	*TSSC4*	Biallelic (LEE *et al.* 1999b)	11p15.5	Tumor suppressing subtransferable candidate 4
rs680	*IGF2*	Imprinted	11p15.5	Insulin-like growth factor 2
rs13390	*PHLDA2*	Imprinted (QIAN *et al.* 1997)	11p15.5	Pleckstrin homology-like domain, family A, member 2
rs1803622	*GAPDH*	Housekeeping	12p13	Glyceraldehydes-3-phosphate dehydrogenase
rs1802710	*DLK1*	Imprinted (KOBAYASHI *et al.* 2000; WYLIE *et al.* 2000)	14q32.2	Delta, Drosophila, homolog-like 1
rs2066707	*ATP10A*	Imprinted (HERZING *et al.* 2001; MEGURO *et al.* 2001)	15q11.2	ATPase, Class V, type 10A (imprinted in brain and lymphoblasts)
rs1053474	*IMPACT*	Biallelic (OKAMURA *et al.* 2000)	18q11.2	Impact homolog; imprinted and ancient
rs12982082 rs8100247	*ZNF331*	Imprinted (PANT *et al.* 2006; POLLARD *et al.* 2008)	19q13.41	Zinc finger protein 331
rs1860565	*PEG3*	Imprinted (MURPHY *et al.* 2001; VAN DEN VEYVER *et al.* 2001)	19q13.43	Paternally expressed gene 3

Table 5: List of human orthologues of mouse imprinted genes.

The SNPs used are listed along the gene official symbol, its chromosomal location, its description and imprinting status according to the Otago website (http://igc.otago.ac.nz/home.html**).**

SNP	Gene	Chr location	Description	Mode of expression in human (Otago imprinted gene catalogue in 2009)
rs1570070 rs998075	*IGF2R*	6q25.3	Insulin-like growth factor 2 receptor	Monoallelic in 3/8 placentas (MONK *et al.* 2006)
rs694812 rs624249 rs316003	*SLC22A2*	6q26	Organic cation transporter	Polymorphic (5/18 monoallelic in placenta)(MONK *et al.* 2006)
rs2076828 rs2292334	*SLC22A3*	6q26	Organic cation transporter	Imprinted in first trimester placenta (MONK *et al.* 2006)
rs1801197 rs2301680	*CALCR*	7q21	Calcitonin receptor	Provisional data (monoallelic in 4/5 brains)(MONK *et al.* 2008)
rs6954345 rs7493	*PON2*	7q21	Paraoxonase	Biallelic in 4 term placentas (MONK *et al.* 2008)
rs1790345	*DHCR7*	11p13.4	7-dehydrocholesterol reductase	Two placental samples (biallelic)(SCHULZ *et al.* 2006)
rs3741041	*AMPD3*	11p15.4	AMP deaminase 3	One placental sample (biallelic)(SCHULZ *et al.* 2006)

SNP	Gene	Chr location	Description	Imprinting in human (Otago imprinted gene catalogue in 2009)
rs6356	*TH*	11p15.5	Tyrosine hydroxylase	No reports
rs1145086 rs1049508	*GATM*	15q21	Glycine amidinotransferase	Biallelic (MONK *et al.* 2008)
rs11855231 rs1562008	*RASGRF1*	15q21	Guanine nucleotide exchange factor	No reports
rs3764574 rs8103779 rs1027392	*USP29*	19q13	Ubiquitin-specific protease	No reports
rs4801433 rs7251328	*ZIM3*	19q13	Zinc finger protein 3	No reports
rs10414299 rs917340	*ZNF264*	19q13	Zinc finger protein 264	No reports

2.6.2.3. Mouse Candidates.

The algorithm in Luedi et al. paper predicted 600 mouse genes, out of a total of 23,788 annotated autosomal genes, to be imprinted (2.5%) and 64% of these candidate imprinted genes were predicted to exhibit maternal expression (LUEDI *et al.* 2005). An old ENSEMBL version 16.30 was used for this study and some gene IDs couldn't be tracked. Some genes had no human orthologue. Only 283 of the 600 mouse candidates had traceable ENSEMBL IDs, human orthologues, and were reported to be expressed in the human placenta.

On the 283 genes remaining, 180 had one or more SNPs fulfilling our location and frequency criteria. The primer design was successful for 100 of them with the Sequenom Assay Design Software (Figure 10 and list in Appendices section) and for 126 of them on the Illumina array.

2.6.2.4. Human Candidates.

The algorithm in Seoighe at al. paper predicted 60 human imprinted candidates (SEOIGHE *et al.* 2006). Of their 60 candidates, five genes were known imprinted genes. Thirty-nine of the remaining 55 candidates were expressed in the placenta and had SNPs located in exons or UTRs with a MAF>15% in the Caucasian population. Three of these genes were also predicted to be imprinted in the mouse. (LUEDI *et al.* 2005): *CTSD, SERPINB2*, and *TGFBI.* In total, primers were successfully designed for 26 of these genes with the Sequenom Assay Design Software and nine of these on the Illumina array (Table 1 of Appendices section).

2.6.2.5. Summary of the successive steps used to select candidate genes for the Sequenom platform.

In order to be tested, the candidate genes had first to be expressed in the human placenta and second to contain informative SNPs (Figure 10).

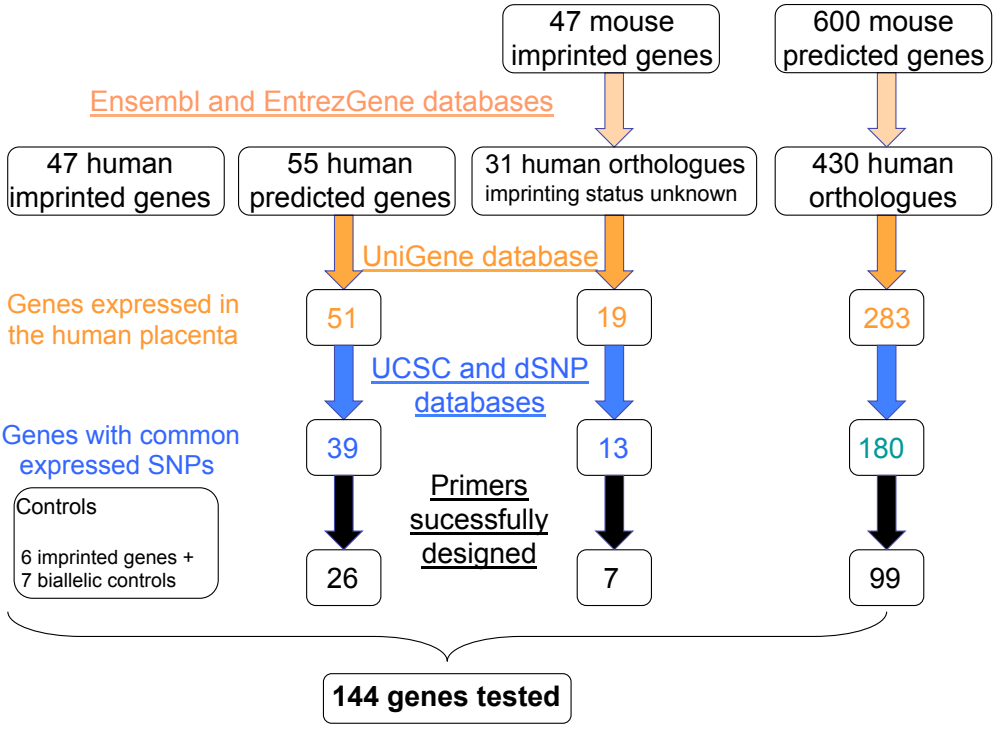

Figure 10: Summary of successive steps to select candidates for Sequenom platform. Bioinformatics tools were used to find human orthologues of murine genes, choose candidate genes expressed in the placenta, find expressed SNPs with a MAF >15% and design suitable primers for the Sequenom assay.

2.6.2.6. Additional candidate genes for the Illumina platform.

2.6.2.6.1. *Genes expressed in the human placenta.*

Sood et al. have studied expression patterns in amnion, chorion, umbilical cord and sections of villus parenchyma of full-term normal pregnancies (SOOD *et al.* 2006). They have found 300 genes that are placenta specific and strongly expressed. The design of primers for the Illumina GoldenGate® Assay was successful for ten placenta specific genes. Sixty-three of the 300 genes were

differentially expressed genes according to the birth weight (SOOD *et al.* 2006). Hence, it was postulated that these genes could be important for the function of the placenta and could potentially play a role in fetal growth. The design of primers for the Illumina GoldenGate® Assay was successful for 35 of these 63 genes.

2.6.2.6.2. *Polycomb genes.*

Polycomb group (PcG) proteins have been discovered in *Drosophilia melanogaster* (LEWIS 1978) and are involved in homeotic (Hox) genes and chromatin regulation. They are highly conserved in the animal kingdom and also in plants. In the latter, one polycomb gene (*MEDEA*) was found to be imprinted and to regulate its own expression (BAROUX *et al.* 2006). It was postulated that human polycomb genes could be good imprinting candidates in human. The words 'polycomb' and 'gene' were used to search the Entrez Gene database (http://www.ncbi.nlm.nih.gov/sites/entrez) and 30 human polycomb genes or genes shown to interact with polycomb proteins were found. 14 of them had exploitable SNPs and the design of primers was successful for six of them (Table 6).

Table 6: List of polycomb genes and polycomb related genes targeted on the Illumina array.

SNP	Gene	Chr location	Full name
rs2295764	*ASXL1*	20q11.1	Additional sex combs like 1
rs1045480	*CTBP1*	4p16	C-terminal binding protein 1, interacts with polycomb group protein
rs17479770	*CUL3*	2q36.2	Cullin 3, ubiquitinates the polycomb protein BMI1
rs1049925	*PHC1*	12p13	Polyhomeotic homolog 1
rs11061	*PHC2*	1p34.3	Polyhomeotic homolog 2
rs1056567	*PHF19*	9q33.2	PHD finger protein 19, overexpressed in many types of cancers

The SNPs used are listed along the gene official symbol, its chromosomal location, its full name and summary of function according to Entrez Gene database (http://www.ncbi.nlm.nih.gov/sites/entrez).

2.7. Mass spectrometry genotyping- Sequenom platform.

2.7.1. MassArray homogeneous MassEXTEND (hME) assay.

The hME assay (Sequenom, Inc.) consists of a primer extension assay for genotyping and quantitation of alleles by MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight) mass spectrometry (Stanssens 2004) in a 384-well microplate. The first step is a PCR amplification of a short product ~100 bp. A 10-mer tag (5'-ACGTTGGATG-3') that is referred to as hME-10 is added by the software at the end of each PCR primer to ease the distinction between unincorporated primers and analytical peaks (the extension also makes the PCR more efficient after the first cycle). After amplification, the PCR products are submitted to a shrimp alkaline phosphatase (SAP) treatment to dephosphorylate and degrade any residual primer. After inactivation of the enzyme, the extension reaction takes place and generates two elongated primers with a different mass for the two alleles. The assay design software selects dideoxy-trinucleotides to be added during the extension reaction in order to maximise this mass difference between the two alleles (Figure 11). A proprietary resin (Sequenom, Inc.) is added to the samples to clean them from salts. Samples are robotically spotted on a microchip (SpectroCHIP, Sequenom) with the MassArray nanodispenser and analysed by SpectroREADER mass spectrometer (Sequenom). It uses laser energy to ionize the samples. The mass to charge ratio is directly analysed in the Mass Spectrometer. Finally, the genotypes are called in real time by the SpectroTyper software (v2.0). The peak area represents the quantity of each primer extension product.

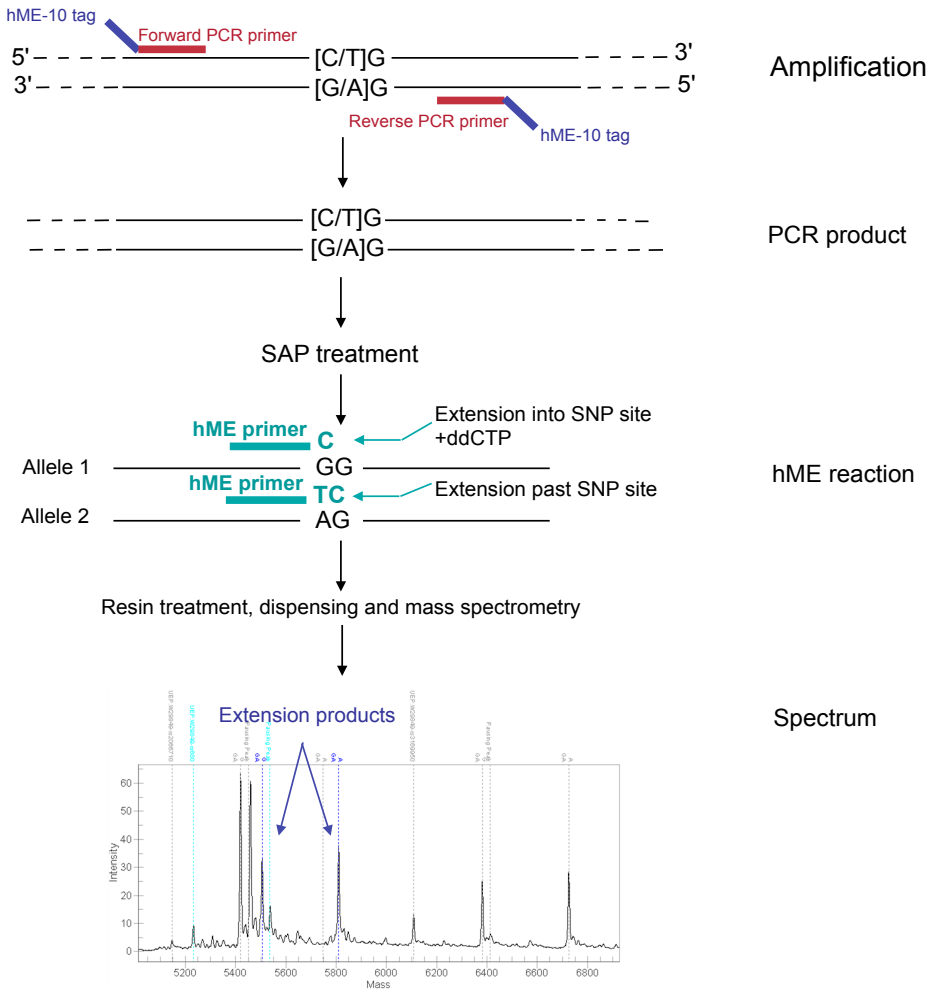

Figure 11: Multiplexed homogeneous MassEXTEND Assay.
This figure depicts a single assay.

Genomic DNA and cDNA are amplified with the same set of PCR primers. The extension reaction incorporates a dideoxy-trinucleotide and the reaction terminates. This extension reaction generates products of a different mass.

2.7.2. Primer design.

The primers were designed using Extend Primer Assay Design v2.0 (Sequenom, Inc.) by Jilur Ghori, (based at Sanger Centre). Three different primers (two for amplification and one allele-specific MassEXTEND primer) were designed for each targeted SNP within the exon or the UTRs of interest. The output file lists all designed primers (3 per SNP), the appropriate termination mix, and the mass of the expected products for each allele (see Appendices section for full list). The primers were designed within the exons or the UTRs so that the same sets of primers could be used on gDNA and cDNA. The reactions were multiplexed in set of threes to reduce costs. Each set of primers was first mapped in in silico PCR to verify that the full length of both primers was within the exon or within the UTR (http://genome.ucsc.edu/cgi-bin/hgPcr?db=hg17). When the primers did not fit within the exon because the SNP was too close to the exon-intron boundary, the design was adapted for a smaller PCR product or for another SNP if there was another one that met the MAF >0.15 criteria for that gene. PCR products length ranged from 78 to 135 bp. A total of 144 primer sets were successfully designed and were synthesised externally by Sigma-Genosys (Haverhill, UK). The full list of primers and termination mixes is listed in Appendix 2).

2.7.3. Input material.

Samples were dispensed robotically (2 µl per well) from the manually made stock 96-well plate into 384-well plates using a Multimek™ 96/384-Channel Automated Pipette (Beckman Coulter). Parental gDNA and placental gDNA were assayed in parallel with the placental cDNA. 4 ng of gDNAs (2 ng/µl) and 2 µl of 1/50 fold dilution of reverse transcription product for the cDNA were used as input material (see section 2.2, 2.3, and 2.5 for samples preparation).

2.7.4. Multiplexed PCR reaction.

First, PCR primers were mixed together in water resulting in a stock solution of 100 µM of each primer. The primer working dilution was then made (375 nM of each primer). Second, a PCR master mix sufficient for the number of planned reactions was prepared. For a 1X reaction mix:

Component	Volume
10X PCR buffer	0.75 µl
dNTPs mix (25 mM)	0.2 µl
Titanium Taq polymerase (Clontech)	0.04 µl
H_2O	0.01 µl

One volume of each primer mix and two volumes of PCR master mix were then mixed and 3 µl/well of this mixture were spread on each 'quadrant' of the 384-well plate. A total of four different primer mixtures were used for each 384-well plate.

The amplification was performed in a MJ Thermocycler with the following cycling profile: activation of the Hot-start Taq at 95ºC for 15 minutes, followed by 45 cycles consisting of: 95ºC for 20 seconds, 56ºC for 30 seconds, 72ºC for 1 minute, and finally an extension cycle at 72ºC for 3 minutes. Plates were centrifuged at 1000 rpm for 1 minute and shrimp alkaline phosphatase (SAP) (Amersham Biosciences) treatment followed. Two µl of the following 1X reaction mix were spread onto the plate.

Component	Volume
10X Thermosequenase buffer	0.2 µl
SAP (1 U/µl)	0.3 µl
H_2O	1.5 µl

The mixture was incubated at 37ºC for 20 minutes, followed by denaturation of the enzyme at 85ºC for 5 minutes. The plates were finally centrifuged at 1000 rpm for 1 minute.

2.7.5. hME reaction.

The extend primers were mixed together for each 100 µM primer in the stock solution. The working primer solution was a 10-fold dilution of the stock solution (10µM of each primer). The hME reaction mix was prepared as follows with Thermosequenase DNA polymerase (Amersham Biosciences):

Component	Volume
Buffer	0.2 µl
ThermoSequenase™ DNA polymerase (32 U/µl)	0.018 µl
H_2O	0.382 µl

Primer mixes and appropriate termination mix (each appropriate ddNTP or dNTP 500 µM) were added to the hME reaction cocktail as follows:

Component	Volume
Extend primer mix	0.5 µl
hME reaction mix	0.6 µl
Termination mix	0.9 µl

Two µl of the final mix were spread onto the plate. The extension reaction was performed with the following cycling conditions: 94ºC for 2 minutes, followed by 55 cycles consisting of: 94ºC for 5 seconds, 52ºC for 5 seconds and 72ºC for 5 seconds. Plates were centrifuged at 1000 rpm for 1 minute.

2.7.6. Desalting and dispensing of samples.

A cationic exchange resin (Clean Resin, Sequenom) was used to clean up samples from salts (Na+, K+, Mg+,…) in order to decrease background noise. Sixteen µl of water and 6 mg of resin were added to the samples. The plates were sealed and rotated for 10 minutes, then centrifuged at 4000 rpm for 4 minutes. Reactions were spotted onto the pad of a 384-well SpectroCHIP (Sequenom) using a SpectroPoint nanoliter sample-dispensing instrument (Sequenom). Chips were then loaded into the mass spectrometer (MassARRAY Analyzer, Sequenom) and the genotype calls were generated in real time by the MassARRAY Typer v3.0.1 software.

2.7.7. Sequenom analysis.

Relative quantitation of alleles was normalised by dividing the peak area value for one allele by the sum of the peak areas for both alleles (A=a/a+b and B=b/a+b). To find new imprinted genes or ASE, genotype calls were filtered to include only the genotypes that had been called with the "conservative" rating. The percentage of genotyping assays called in this way for each SNP was referred to as the *success rate* (SR) and was calculated for gDNA and cDNA. The ratio of cDNA to gDNA SR (SRratio) was used to filter out lowly expressed genes. Genotyping with a SRratio $\geq$ 75% was taken forward in the analysis. Calls were then filtered to select trios with heterozygous placental genomic DNA. The statistical analysis (t-test) was carried out in R (R Foundation for Statistical Computing, Vienna, Austria) by Ian Sudbery, Sanger Centre. On these trios, a one-tailed paired t-test was used, for each SNP, to compare allelic quantification of the two alleles in placental cDNA and in placental genomic DNA. P-values were then adjusted using the Benjamini-Hochberg method to control false discovery rate (BENJAMINI and HOCHBERG 1995).

2.8. Technology Illumina allele specific expression (ASE) array.

2.8.1. Overview of the technology.

The platform has a 96-array format, allowing the examination of 96 samples per BeadChip. Each array contains 1,536 different bead types; each bead type interrogates a different SNP (Figure 12). There is an average of 30 copies of each bead type per array (GUNDERSON *et al.* 2005). Each bead type has ~700,000 copies of a particular oligonucleotide probe covalently attached to it, allowing reliable quantification of expression (KUHN *et al.* 2004) (Figure 13).

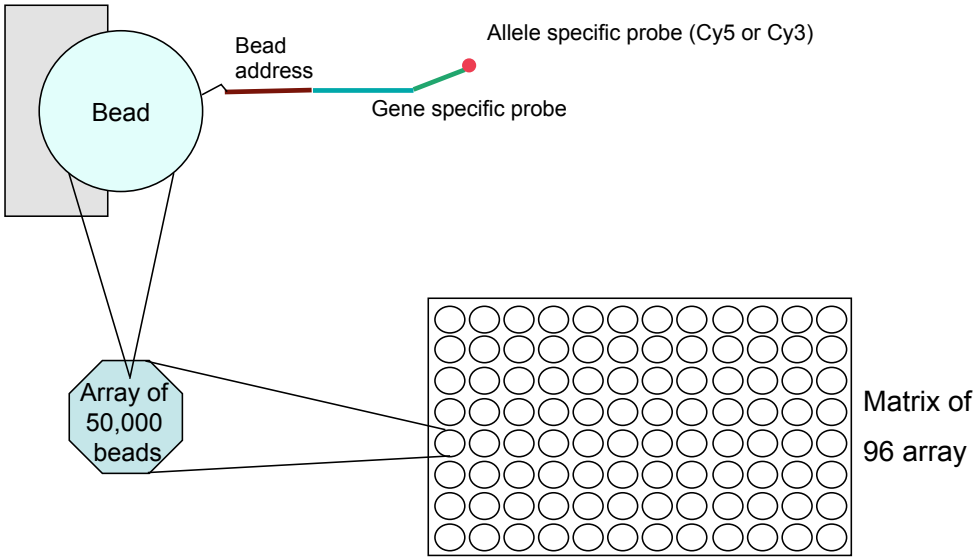

Figure 12: Illumina technology adapted from Kuhn et al. (KUHN *et al.* 2004).

For each placental sample, gDNA and cDNA were run on the same plate. The experiment was repeated on a different day. Parental genotyping was performed on a separate plate and was not replicated. The arrays with a low dynamic range were discarded and repeated.

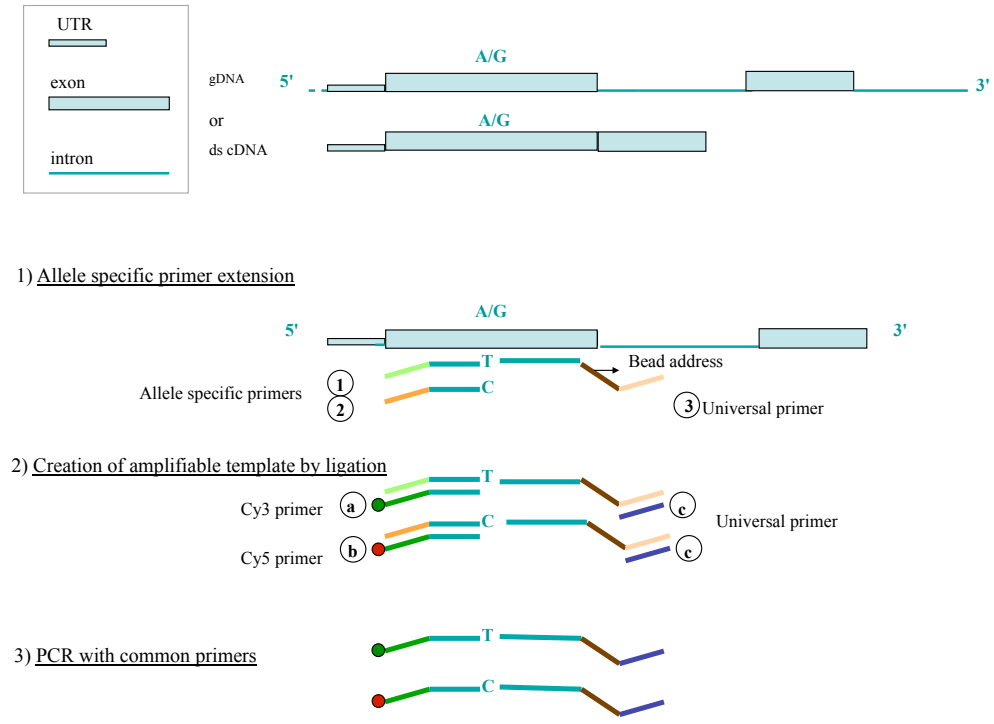

Figure 13: Illumina technology overview.

Genomic DNA was normalised to 50ng/µl and a total of 250 ng was used. Double stranded cDNA was synthesised and RT product made from 250 ng of RNA is used per array. Three oligonucleotides (oligos) are designed for each SNP. Two are allele specific oligos (1 and 2) and one is locus specific (3). Each has a *universal* primer sequence for PCR (a, b and c) and the locus specific primer also contains an *address* sequence that will be used to attach to the bead. After extension and ligation, the ligated primers are amplified by PCR using Cy3 and Cy5 labelled allele-specific 5' primers and a universal 3' primer. The PCR products are then hybridised to a Sentrix Array Matrix (SAM) containing the 1536 beads covered by an oligo complementary to the bead address. The genotype of each sample can be determined according to the colour distribution. Homozygotes have only one colour and heterozygotes have two.

2.8.2. Candidate genes targeted.

The oligo pool of 1536 SNPs of the GoldenGate ASE Array (Illumina, Inc., USA) included 18 known imprinted genes, four housekeeping genes, 11 genes shown to be preferentially expressed in the placenta (SOOD *et al.* 2006), ten genes predicted to be imprinted in humans (SEOIGHE *et al.* 2006), ten orthologues of mouse imprinted genes, 35 genes that are differentially expressed according to infant weight (SOOD *et al.* 2006), six polycomb genes and 124 human orthologues of genes predicted to be imprinted in mouse (LUEDI *et al.* 2005); all of which were expressed in the placenta according to the Unigene database (http://www.ncbi.nlm.nih.gov/UniGene) (see Supplementary Table 1 for list of SNPs and genes). All SNPs chosen were located within the exons or UTRs of the targeted genes in order to be present in the spliced mRNA and to be able to compare gDNA and cDNA hybridisations. SNPs with the highest MAF in our population in the single nucleotide polymorphisms database (dbSNP Build ID: 125 and 126, http://www.ncbi.nlm.nih.gov/SNP/) were chosen. Best Illumina design scores in our candidate genes were preferred: SNPs less than 40 bp from the exon-intron boundaries were discarded; SNPs with a Score >0.6 had already been tested on the platform and had a success rate of 90-95%; the assays with a score of 1.1 had already been successfully run by Illumina. The last selection on this list was whether two SNPs could be interrogated per gene while linkage disequilibrium (high r^2) avoided where possible.

The 1536 exonic SNPs of this Illumina® ASE Bead Array™ targeted 932 genes (Appendix 1 Table 2). 357 SNPs of the 1536 SNPs on the array were of direct interest to my work. These SNPs targeted 214 genes.

2.8.3. Hybridisation of the Illumina GoldenGate protocol.

Matthew Forrest and the members of the Genotyping Facility Team at the Sanger Institute carried out the hybridisation steps of the Illumina GoldenGate protocol. Paired gDNA (250ng) and double-stranded cDNA (made from 250 ng total

RNA, see above) were identically processed and hybridised to a standard 96-sample Sentrix Array Matrix according to the manufacturer's instructions for GoldenGate genotyping assays (Illumina, Inc., USA) (FAN *et al.* 2006). After a 16 hour-hybridization and washing, each array was scanned by Illumina BeadScan software (Illumina, Inc., USA) to produce 2 TIFF images (one for Cy3 and one for Cy5). For each placental sample, gDNA and cDNA were assayed on the same plate, and the whole plate analysis was replicated on a different day. For two cDNA replicates (on 48 in total), cDNA amplification was not obtained. Parental gDNA genotyping was performed on a separate plate and not replicated. The genotypes were called using Illumina's proprietary software (BeadStudio) (Illumina, Inc.) and manually curated (Figure 14).

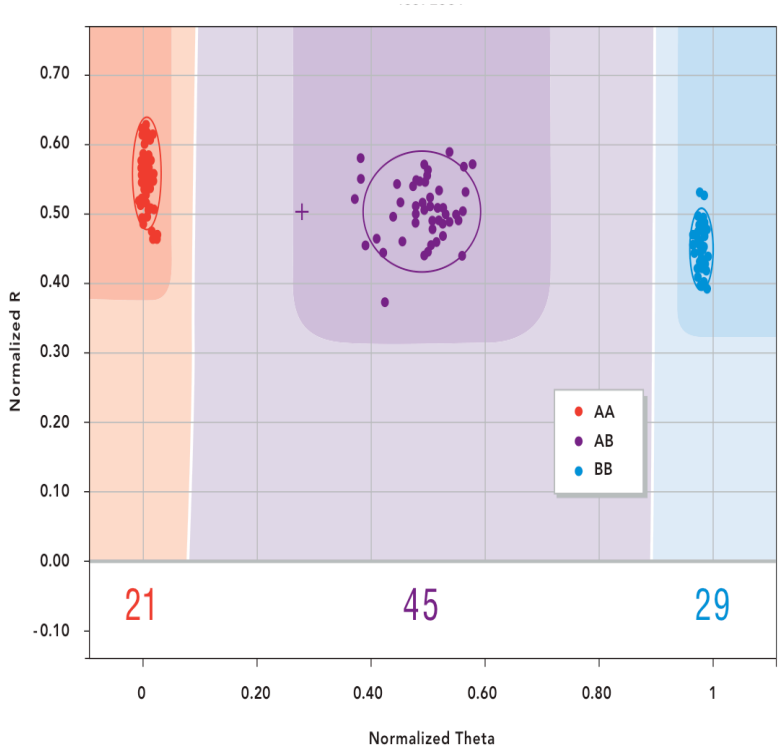

Figure 14: Example of manual clustering of genotypes using BeadStudio software (Illumina, Inc.)

The composition of the trios was imported so that Mendelian errors could be highlighted during the manual curation of the genotyping. Arrays with a low dynamic range were discarded and repeated.

2.8.4. Quality control and description of the normalization method.

A collaboration with Matthew Ritchie from Prof. Simon Tavaré's group at CRUK Cambridge Research Institute was used for the normalisation and subsequent analysis of the microarray data. Matthew Ritchie performed the normalization of data, so that I could analyse and produce graphs for all known imprinted genes and then candidate genes.

First, the raw Cy3 and Cy5 intensities from all beads on an array were examined and arrays with hybridisations of poor quality were filtered out. Intensities were then quantile normalised between channels. Log-ratios ($\log_2$(Cy5/Cy3)) and average log-intensities ($^1/_2\log_2$(Cy5*Cy3)) were calculated for each bead on each array. Outliers greater than 3 mean absolute deviations (MADs) from the median of each bead type were removed as per Illumina's standard method and the remaining values were averaged to obtain a summary log-ratio and average log-intensity for each bead type on each array (there is a mean of ~30 beads per SNP tested). The summarized data were normalized per array by median centering the log-ratios to have a median of zero on each array. The intensity distribution of ds cDNA was lower than for gDNA (i.e. ds cDNA yielded lower signal intensity), and between-array normalization was performed for ds cDNA and gDNA separately (RITCHIE *et al.* 2010).

2.8.5. Experimental Data analysis.

After data normalisation, all downstream analysis has been performed on the log(Red/Green) scale. 141,312 genotypes in total have been called across the cohort. The data were first manually analysed one candidate SNP at a time by me. To be informative, the placental gDNA of a trio needed to be heterozygous with at least on of the parents homozygous for the SNP of interest. All informative samples were used to produce graphs for each SNP.

To test for ASE, Matthew Ritchie used the following method. Linear models were fitted to the cDNA log-ratios to summarise the replicate observations. After empirical Bayes shrinkage of the SNP-wise variances, moderated *t*-statistics were calculated (SMYTH 2004). Raw *p*-values from these *t*-tests were adjusted globally for multiple testing using the method of Benjamini and Hochberg to control the false discovery rate (BENJAMINI and HOCHBERG 1995). In addition, our criteria for ASE required that SNPs satisfy the following conditions: (1) average intensity across all samples greater than 11.25 (Illumina arbitrary fluorescence units); (2) at least 80% of homozygotes with adjusted *p*-values less than 0.01 and absolute log-fold-changes greater than 0.585 and (3) at least 2 heterozygotes (based on BeadStudio calls from gDNA samples) with adjusted *p*-values less than 0.01 and absolute log-fold-changes greater than 0.585. The intensity cut-off was based on the concordance between Illumina and Sequenom data (see below and section 4.3.2. of Chapter 4), with probes expressed below this level less reliably quantified on the Illumina arrays. The log-fold-change cut-off of 0.585 was based on the mixture data (see below). This experiment showed that true positives were more difficult to detect on the Illumina arrays in mixtures at or below 60:40/40:60 (equivalent to absolute log-ratios less than $\log_2(60/40) = 0.585$). The homozygote criteria (2) ensured that the two alleles could be reliably distinguished in the cDNA samples. All analyses were carried out by Matthew Ritchie in R using the beadarray (DUNNING *et al.* 2007) and limma packages.

2.8.6. Imprinted genes spectrum of silencing.

The spectrum of allelic silencing in imprinted genes was examined for all controls on the array. A graph (Figure 29, 31, 35, 37, 39, and 41 in Chapter 4) representing the mean proportion of the silenced allele across all informative individuals and its corresponding standard error (+/- one standard error) was produced. The raw allelic values were averaged for the expressed (either paternally or maternally inherited) and for the silenced alleles across all informative samples for a particular SNP. The means and standard errors were calculated on the logit scale for each heterozygote (i.e. the signal from each

heterozygote was transformed as log(p/1-p), where p = silenced/(silenced allele + expressed allele), and then for all heterozygotes. The results were converted back to the original scale.

2.8.7. Control gDNA mixture data created from HapMap individuals.

For the control experiment, gDNA mixtures of two HapMap individuals (NA12892:NA19092) (Coriell, Camden, New Jersey, United States gifted from Susana Campino at the Sanger Institute) were created by Matthew Forrest in the following proportions: 0%:100%, 5%:95%, 91%:9%, 83%:17%, 67%:33%, 64%:36%, 60%:40%, 56%:44%, 50%:50%, 44%:56%, 40%:60%, 36%:64%, 33%:67%, 17%:83%, 9%:91%, 5%:95% and 100%:0%. Matthew Forrest and the Genotyping Facility team at the Sanger Institute hybridized each mixture in duplicate using the same experimental protocol. Data were pre-processed, normalised as described in the previous section, and analysed by Matthew Ritchie (RITCHIE *et al.* 2010).

Briefly, a linear model was fitted to each SNP as described previously, and contrasts were obtained to give all pairwise comparisons between a given mixture and the 50%:50% mixture. This corrects for dye biases and systematic shifts which are present for SNPs which are either heterozygous and homozygous (i.e. AA:AB, BB:AB, AB:AA or AB:BB) or have the same genotype (AA:AA, BB:BB or AB:AB) in the two individuals. Moderated *t*-statistics were calculated using the empirical Bayes shrinkage procedure (SMYTH 2004) to test the null hypothesis that each contrast was equal to 0 (i.e. no allelic imbalance). Sensitivity and specificity calculations were made for each contrast by ranking SNPs by their log-odds and using *a priori* genotype information on which SNPs are true positives/negatives for allelic imbalance.

Genotypes for NA12892 and NA19092 were downloaded from HapMart (http://hapmart.hapmap.org/BioMart/martview, version 21, NCBI Build 35) for

the SNPs present on the array. SNPs with known allelic imbalances between these individuals (782), such as those which are either homozygous and different (AA:BB or BB:AA), or heterozygous and homozygous (AA:AB, BB:AB, AB:AA or AB:BB), form the true positive set. SNPs which have the same genotype for each individual (AA:AA, BB:BB or AB:AB) should not change with mixing concentration, and comprise the true negative set (533). SNPs with missing data (15 with NN calls) and those with IDs that could not be found in HapMart (206v21) were excluded from the analysis.

2.8.8. Mendelian errors in the experimental samples.

The composition of the trios was imported so that Mendelian errors could be highlighted during the manual curation of the genotyping. This enabled a more accurate call of the genotypes. One family displayed many Mendelian errors. After inspection of the genotypes for each SNP, it was concluded that parents were most likely not the baby's genetic parents. A mislabelling of the tubes at sampling was the most likely explanation as previous experiments in the laboratory also corroborated this finding (Abu-Amero S, personal communication).

2.8.9. Illumina and Sequenom platforms correlation.

The genotyping results obtained for the SNPs tested on the same samples on both platforms were compared. Pearson correlation coefficients were calculated by Matthew Ritchie for 38 SNPs using log-ratios (log-ratios calculated as $\log_2$[(Sequenom allelic ratio_x+1)/(Sequenom allelic ratio_y+1)]). An arbitrary intensity threshold of 11.25 units (average $\log_2$ fluorescence) was determined as below this the correlation was weaker (see section 4.3. of Chapter 4).

2.9. Polymerase chain reaction (PCR) amplification and Sanger sequencing.

2.9.1. PCR.

Nucleic acid sequence was exponentially amplified by PCR (MULLIS *et al.* 1986; SAIKI *et al.* 1986; MULLIS and FALOONA 1987). Double-stranded DNA or single-stranded cDNA were denatured in the presence of two oligonucleotide primers, four deoxyribonucleoside triphosphates (dNTPs) and DNA polymerase. The primers were complementary to the sequence of interest. The PCR reactions were carried out in a 96-well plate (ABgene) or 0.5 ml microcentrifuge tubes in a *DYAD Peltier* Thermal Cycler (MJ Research, *UK*). The lids of the thermocycler were heated (100ºC) during the PCR cycle to minimize evaporation.

For a standard PCR reaction, DNA polymerase (Biotaq™, Bioline) was used with each dNTP mixed in equal amount (Promega).

Component	Volume
DNA or cDNA template	50 ng (1µl)
Forward primer	100 ng (0.5µl)
Reverse primer	100 ng (0.5 µl)
dNTPs	10 mM (0.5µl)
(NH4)2SO4 buffer (10x)	5 µl
$MgCl_2$	1.5 mM (0.75 µl)
Betaine (5 M)	7.5 µl
Taq polymerase	1 unit (0.2 µl)
MilliQ water	adjust to 25 µl

The cycling conditions were as follows:

Step	Temperature	Time
1) Initial denaturation	96ºC	5 minutes
2) Cycle denaturation	94ºC	30 seconds
3) Annealing	53ºC→63ºC	30 seconds
4) Extension	72ºC	30 seconds
5) Cycling from step 2 to step 4 for 30 to 45 cycles		
6) Final extension	72ºC	5 minutes

The annealing temperature was adapted for each primer pair according to the gradient profile results performed during the set-up stage of each PCR reaction The extension time was increased when the PCR product was longer than 500 bp (typically by 30 seconds per 500 bp).

2.9.2. PCR/ sequencing primers.

Primers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/) (ROZEN and SKALETSKY 2000) to anneal to the sequence upstream and downstream of the SNP or endonuclease restriction site of interest. Sequences for the targeted genes and SNPs were obtained from the reference sequence of the human genome using the UCSC genome browser (http://genome.ucsc.edu/). In Primer3, the human mispriming library was chosen to exclude repetitive regions and the sequence of the locus of interest submitted. The candidate SNP was labelled as the target and other SNPs in the sequence were labelled as excluded regions. The product size and primer size (between 18 and 27 bp with an optimum around 20 bp) were specified. The table of thermodynamic parameters of Breslauer et al. was used by the webtool to predict the DNA duplex stability from the base sequence (BRESLAUER *et al.* 1986). All primer sequences used are listed in Appendix 3.

All primers were synthesised externally by Eurofins MWG/Operon (London). Primers were provided dried down with the yield supplied in µg. Stock solutions

of primers (1µg/µl) were stored at -20ºC and working dilutions prepared at 100 ng/µl. The primer pairs were tested at different temperatures (gradient from 53 to 63ºC) using the standard cycling conditions on the thermal cyclers to determine the optimal PCR conditions.

2.9.3. Agarose gel electrophoresis.

The size and integrity of PCR amplicons and restriction endonuclease digestions (see section 2.10.3. below) were analysed using size separation by agarose gel electrophoresis, staining with ethidium bromide (EtBr) and visualisation with UV light. The Agarose concentration used was 1% for PCR and 3% for restriction digests. One percent gels were made with 1 g of agarose (Invitrogen) in 100 ml Tris acetate EDTA (TAE) buffer (3% gels with 3 g of agarose). Agarose was brought to the boil in a microwave oven in 1x TAE buffer to be dissolved and 1 µl of EtBr stock (10 mg/ml) per 100 ml was added when the solution had cooled down to approximately 50ºC. Gels were set in mini casts. Samples were prepared for electrophoresis by adding 3 µl bromophenol blue loading dye (Promega) diluted 1:4 in 80% glycerol, to 8 µl PCR product or 17 µl restriction digest. Five µl of 100 bp DNA marker (Promega), loaded in parallel, allowed estimation of fragment size. Loaded gels were electrophoresed in 1x TAE buffer at 120 Volts for 30 minutes to one hour in a gel electrophoresis apparatus. The DNA/RNA bands were visualised using a UV transilluminator (ImageMaster, Amersham Pharmacia Biotech) equipped with a Fuji film thermal imaging system. The pictures were printed or photographed or stored as digital images.

2.9.4. Sanger sequencing.

2.9.4.1. PCR and RT-PCR products 'clean-up' protocol.

Before automated sequencing, PCR products were first purified using microCLEAN (Microzone, Cambio, UK) to remove unincorporated primers and

excess dNTPs. Five µl of PCR template were thoroughly mixed with an equal amount of microCLEAN reagent. DNA was precipitated by centrifugation at 3000 rpm for 40 minutes. The liquid phase was removed and sequencing reaction followed.

2.9.4.2. Fluorescent-labelled (dye terminator) cycle sequencing.

Sequencing was carried out using the BigDye® Terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) and 0.1 µl of primer (100 ng/µl) according to this reaction protocol:

Component	Volume
BDT v1.1	0.5 µl
Buffer	1.5 µl
Q solution (Qiagen)	2 µl
Primer (100 ng/µl)	0.1 µl

The following cycle sequencing programme was then used:

Step	Temperature	Time
1) Initial denaturation	96ºC	1 minute
2) Rapid thermal ramping to	96ºC	30 seconds
3) Rapid thermal ramping to	53ºC	15 seconds
4) Rapid thermal ramping to	60ºC	4 minutes
5) Steps 2, 3 and 4 are repeated another 29 times		
6) Rapid thermal ramping to	20ºC	2 minutes

The content of the plate (or tubes) was briefly spun down before precipitation.

2.9.4.3. Sequences precipitation.

170 µl of 'precipitation solution' (5 ml 3M NaAc, 30 ml 100% EtOH, made up to 250 ml with H_2O) was added to each reaction and plates were centrifuged at 4000 rpm for 40 minutes at room temperature. The supernatant was removed by inverting the plates onto filter pads and spinning at 400 rpm for 30 seconds. The pellets were washed with 170 µl 70% EtOH and spun again for 10 minutes. The wash was removed and the pellet allowed to partially air dry, before being dissolved in 10 µl of 0.1 TE. The sequencing reactions were loaded onto ABI Prism 3730 capillary machines (Applied Biosystems (ABI), CA) and the read-out visualised using Sequencher™ v4.6 (Gene Codes Corporation, MI).

2.10. DNA methylation analysis.

2.10.1. Combined bisulphite restriction analysis (CoBRA).

A quantitative technique called CoBRA has been used to determine DNA methylation levels at specific loci in gDNA (XIONG and LAIRD 1997). The gDNA was first treated with standard sodium bisulphite that converts unmethylated cytosine residues to uracil by selective deamination. Methylated cytosine residues are unaltered by the treatment (FROMMER *et al.* 1992). New restriction enzyme sites are potentially created in the converted sequence and can be exploited to discriminate between methylated and unmethylated CpG sites (SADRI and HORNSBY 1996) (Figure 15).

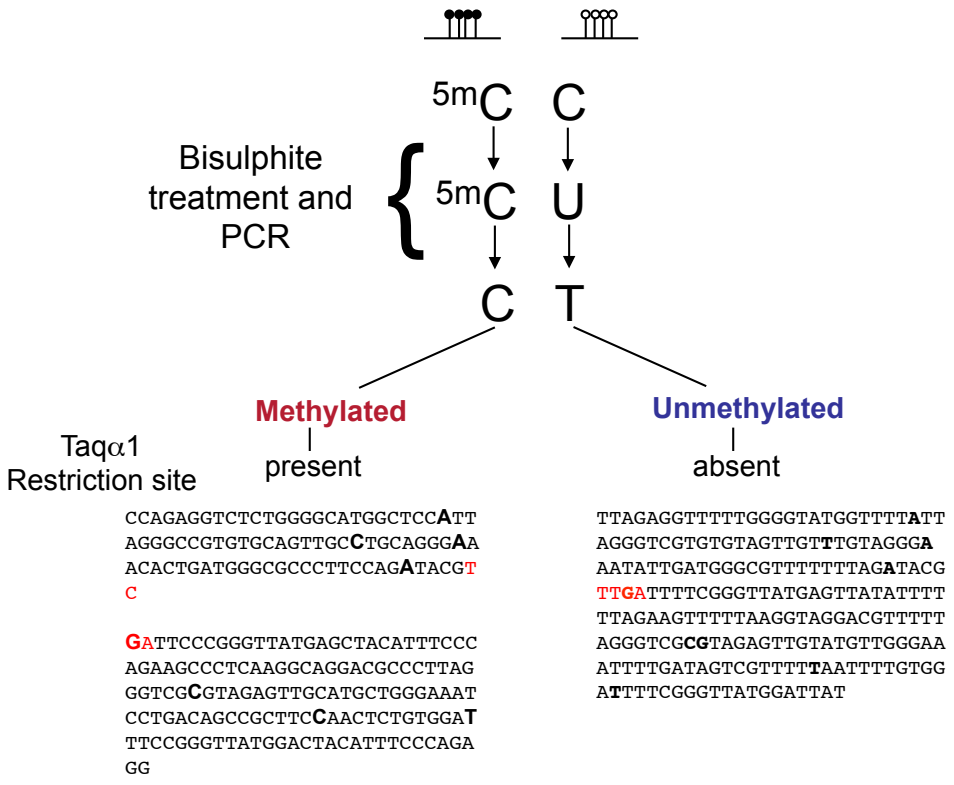

Figure 15: Schematic representation of the combined use of bisulphite treatment and restriction analysis in the CoBRA methylation assay.

Examples of a methylated sequence and of an unmethylated sequence are shown. The restriction enzyme, in this example Taqα1, recognises a CG (sequence in red) site and a digested product obtained when the sequence is methylated.

Bisulphite converted gDNA samples were prepared and cleaned using the EZ DNA methylation-Gold™ kit (Zymo, CA) according to the manufacturer's instructions. The optimal input DNA was found to be 500 ng. Two successive elutions were performed in 10 µl of elution buffer provided in the kit (Zymo, CA) and diluted 10-fold before use.

2.10.2. Bisulphite primers.

For each CpG island of interest, bisulphite primers were designed using the MethPrimer webtool (http://www.urogene.org/methprimer/index1.html) (LI and DAHIYA 2002). The unmodified source sequence was pasted and primers for bisulphite sequencing PCR were chosen using the CpG island prediction setting. The endonuclease restriction site was the 'target' and regions containing SNPs were excluded. In order to be able to determine the parent-of-origin of the methylated or unmethylated sequence, a SNP was also included in the target sequence when possible. Primers were slightly longer than for standard PCR: between 20 and 30 bp and are listed in Appendix 3.

2.10.3. PCR amplification and endonuclease restriction of converted gDNA samples for CoBRA.

Hotstart Taq polymerase (Qiagen, West Sussex, UK) was used for 45 PCR cycles to amplify converted gDNA samples. The amplicons were digested using either Taqα1 (-TCGA-) or Tai1 (-ACGT-) restriction enzymes (New England Biolabs) in a 15 µl reaction.
Restriction reaction was as follows:

Component	Volume
PCR product	3 µl
10X Buffer provided by manufacturer	1.5 µl
Enzyme	1 µl
MilliQ water	9.5 µl

The digestion reaction was incubated for 3 hours at 65ºC for Tai1 enzyme and at 37ºC for Taqα1. Digested products were resolved on 3% agarose gels stained with EtBr (Figure 16).

Figure 16: Schematic representation of a CoBRA experiment.
In lane A, an undigested product acts as a control. In lane B, the product has not been cleaved by the enzyme, which suggests it is unmethylated or methylated depending on the enzyme used. In lane C, the product appears completely digested, which is suggestive of methylation. In lane D, digested and undigested products are present which suggests the presence of both methylated and unmethylated DNA.

2.11. TA cloning and sequencing of bisulphite DNA.

To analyse the methylation pattern of the CpG island of interest more broadly than a single CpG site resolution, PCR products were TA cloned and individual clones were sequenced (ZHANG *et al.* 2009).

2.11.1. Ligation into the plasmids.

One to three µl of crude PCR product was ligated into pGEM®-T ® Vector System (Promega) as per manufacturer's instructions, overnight at 4°C with the 2X Rapid Ligation Buffer and the T4 DNA Ligase supplied with the system.

2.11.2. Transformation using the Vector Ligation Reactions.

Ligations were then incubated at 4°C with JM109 high efficiency competent bacterial cells (Promega) for 30 minutes. The bacterial cells were then heat shocked at 42°C for 45 seconds in a pre-heated water bath and immediately returned on ice for 2 minutes to allow the plasmids to enter by the opened pores of their outer membrane. The successful cloning of an insert into the pGEM®-T Vector should interrupt the coding sequence of ß–galactosidase. Hence, the bacteria that have been successfully transformed should be white and ampicillin-resistant. The bacterial cultures were grown in 100 µl LB broth at 37°C for 30 minutes while being shaken (~150 rpm) and then plated onto LB-agar plates containing ampicillin, IPTG (Isopropyl ß–D-1-thiogalactopyranoside) and X-Gal and incubated overnight at 37°C.

2.11.3. Screening of transformants for inserts.

White colonies were selected for sequencing and resuspended in 100 µl of LB-broth using a pipette tip. The resuspended colonies were incubated at 37ºC for 1

to 2 hours. Two µl of each colony was amplified by standard PCR reaction as outlined in section 2.9 with M13 forward and reverse primers (Appendix 3) or the specific primers designed for the CpG island of interest and cycling conditions as follows:

Step	Temperature	Time
1) Initial denaturation	96ºC	5 minutes
2) Cycle denaturation	94ºC	30 seconds
3) Annealing	56ºC	30 seconds
4) Extension	72ºC	30 seconds
5) Cycling from step 2 to step 4 for 35 cycles		
6) Final extension	72ºC	5 minutes

The integrity and size of PCR products was checked by gel electrophoresis. The PCR products obtained were sequenced as in section 2.9 with the M13 reverse primer or one of the specific primers designed for the CpG island of interest. Sequences were analysed to determine bisulphite conversion of CpG sites using Bisulphite Sequencing DNA Methylation Analysis (BISMA) webtool (http://biochem.jacobs-university.de/BDPC/BISMA/index.php) (ROHDE *et al.* 2008). The unconverted reference sequence for each CpG island studied and the sequencing results were uploaded. BISMA extracts the conversion rate, the percentage of insertions/deletions, and the sequence identity percentage. Sequences complying with the quality control are aligned and BISMA counterchecks for the existence of clonal amplification from the same genomic template. After the data processing, the methylation pattern of sequences is represented in one picture where the clones are sorted according to their methylation level. The blue and yellow colours were chosen to represent the methylated and unmetylated clones respectively. The percentages of unmethylated and methylated CpGs are also calculated by BISMA.

2.12. Table of URLs visited.

Table 7: URLS visited.

ArrayExpress	http://www.ebi.ac.uk/microarray-as/ae/
Catalogues of imprinted genes	http://igc.otago.ac.nz/home.html
	http://www.geneimprint.com
Bisulphite Sequencing Methylation Analysis (BISMA)	http://biochem.jacobs-university.de/BDPC/BISMA/
Bio GPS atlas	http://biogps.gnf.org/
Collaborative Consensus coding sequence	http://www.ncbi.nlm.nih.gov
Copyright	http://www.copyright.com/
dbEST database	http://www.ncbi.nlm.nih.gov/repository/dbEST
dbSNP database	http://www.ncbi.nlm.nih.gov/SNP
Ensembl genome browser	http://www.ensembl.org/index.html
Genevar	http://www.sanger.ac.uk/resources/software/genevar/
HapMart	http://hapmart.hapmap.org/BioMart/martview version 21, NCBI Build 35
Harwell Mouse Imprinting	http://www.mgu.har.mrc.ac.uk
NCBI Gene loci, Genome database, Expression database, Orthologue search	http://www.ncbi.nlm.nih.gov/sites/entrez
NCBI Basic sequence alignment tool	http://www.ncbi.nlm.nih.gov/BLAST
HapMap Project	http://www.hapmap.org/, http://www.sanger.ac.uk/humgen/hapmap3/
HUGO Gene Nomenclature Committee	http://www.genenames.org/
In silico PCR	http://genome.ucsc.edu/cgi-bin/hgPcr?db=hg17
MethBlast	http://medgen.ugent.be/methBLAST/
MethPrimer	http://www.urogene.org/methprimer/index1.html

miRBase	http://www.mirbase.org/
NEBcutter v2.0	http://tools.neb.com/NEBcutter2/index.php
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM
piRNA database	http://pirnabank.ibab.ac.in/
Primer 3	http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
Rfam	http://rfam.sanger.ac.uk/
SymAtlas (Novartis Foundation) (now BioGPS: The Gene Portal Hub)	http://biogps.gnf.org/ - goto=welcome
University of California Santa Cruz (UCSC) genome browser	http://genome.ucsc.edu/

Chapter 3. Sequenom® quantitative genotyping.

3.1. Introduction.

The alleles of imprinted genes are differentially expressed depending on their parent-of-origin. Imprinted or monoallelic expression can be detected by exploitation of polymorphisms in the spliced transcript, for example SNPs occurring within the exon or UTR of any gene. These SNPs are used to quantify the expression of each allele in heterozygous individuals. A cohort of human placental DNA and RNA and parental genomic DNA trio samples was used to screen for imprinting. If the placental genomic DNA is heterozygous for a SNP within the gene under study and if at least one parent is homozygous for this SNP, then the parental origin of the alleles and any type of differential allelic expression, imprinted, allelically biased, or random monoallelic can be detected. The trio is informative (Figure 17, see next page).

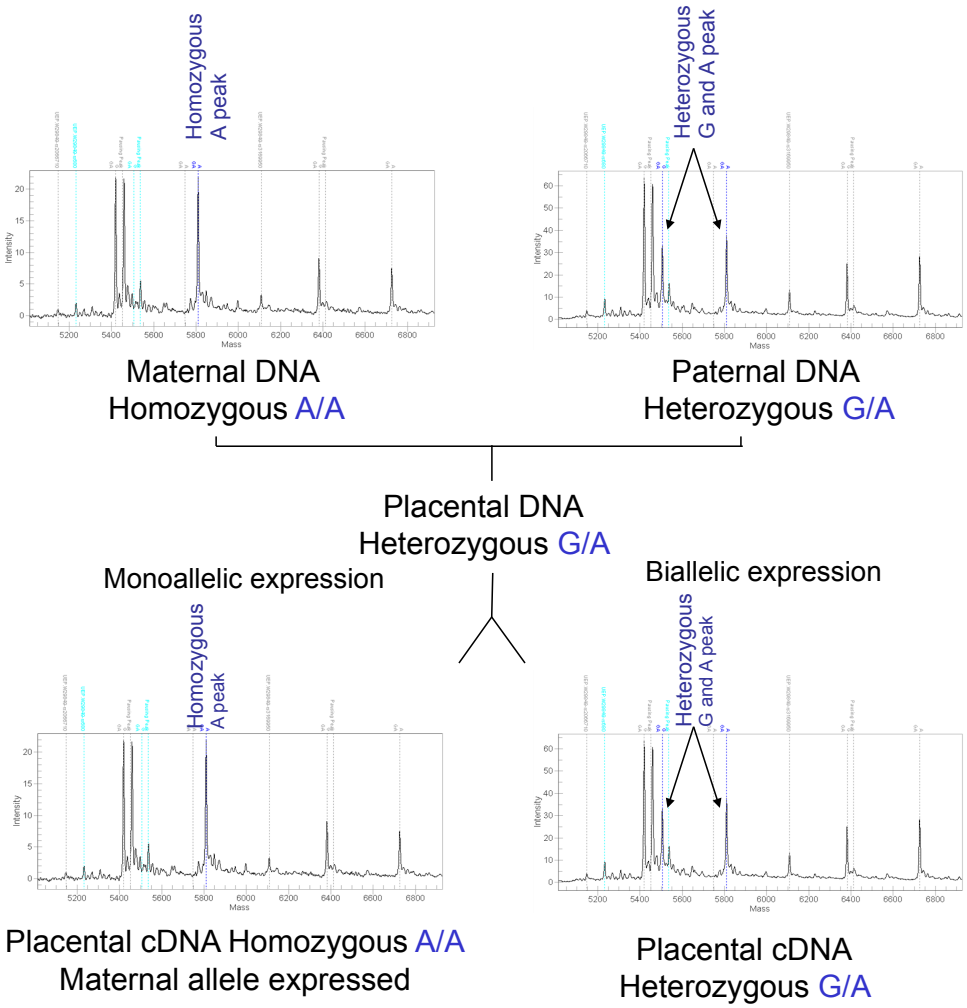

Figure 17: Informative trio and example of Sequenom mass spectrometry allelic quantification.

Heterozygous placental samples are informative. Potential monoallelic expression can be detected in the corresponding cDNA sample and the parental origin of the expressed allele can be determined if at least one parent is homozygous.

3.1.1. Allele detection technology.

Various quantitative methods enable the measurement of the expression of each allele at SNPs and thus the identification of allele specific expression (YAN *et al.* 2002; BRAY *et al.* 2003; LO *et al.* 2003; KNIGHT *et al.* 2004; PASTINEN and HUDSON 2004; PANT *et al.* 2006; GIMELBRANT *et al.* 2007; MILANI *et al.* 2007; POLLARD *et al.* 2008; SERRE *et al.* 2008a).

The MassArray system (Sequenom, Inc.) consists of a primer extension assay for genotyping and quantitation of alleles by MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight) mass spectrometry (STANSSENS *et al.* 2004). The genotypes are called in real time and the peak areas represent the quantity of each primer extension product. This technology has several advantages. First, the assay design is completely flexible. Second, the extension products of several loci can be tested in one reaction (multiplexing). Third, it allows for many samples (384-plate) to be tested simultaneously, only using a small amount of input material, which is important when precious samples are used.

The array has previously been used successfully in noninvasive prenatal screening for trisomy 21 (LO *et al.* 2007; PALOMAKI *et al.* 2011). In this study, it allowed the quantification between alleles of a SNP present in the placenta-specific 4, *PLAC4,* gene, which is transcribed from chromosome 21 and expressed preferentially in placenta. This quantification was accurate enough to detect only a small bias due to an additional chromosome 21 (LO *et al.* 2007).

3.1.2. SNPs.

More and more variations of the genomic sequence at the single base level (SNPs) or insertion/deletion have been mapped over the years (ALTSHULER *et al.* 2000; 2005; INTERNATIONAL HAPMAP 2005; FRAZER *et al.* 2007; INTERNATIONAL HAPMAP *et al.* 2007; ALTSHULER *et al.* 2010). The SNP Mapping Consortium started the catalogue of human sequence variation with

1.42 million SNPs (SACHIDANANDAM *et al.* 2001). Phase I of the International HapMap project has provided approximately 1.3 million SNPs with a MAF > 0.05 (INTERNATIONAL HAPMAP 2005). The availability of the human sequence, the databases of common SNPs, the web-based tools to use them and the high-throughput genotyping techniques have enabled researchers to address genetic associations in common diseases and potential response to pharmacological treatments (ALTSHULER and DALY 2007; BOWCOCK 2007). With the release of phase II of the HapMap project, the number of SNPs was 3.1 million (INTERNATIONAL HAPMAP *et al.* 2007). Recently, the HapMap III has increased further the catalogue of human sequence variations and will be merged with the two first phases of the project (ALTSHULER *et al.* 2010). All these genotypes and their frequencies in the different populations are available in public databases: dbSNP (http://www.ncbi.nlm.nih.gov/SNP), HapMap database (http://www.hapmap.org) and UCSC (http://genome.ucsc.edu/).

The DNA samples for the HapMap were initially from a total of 270 people: 30 trios (two parents and an adult child) of Yoruba (YRI) families of Ibadan (Nigeria), 45 unrelated individuals from the Tokyo area (Japan, JPT), 45 unrelated Han Chinese from Beijing (CHB). The last thirty trios were collected in 1980 from U.S. residents with northern and western European ancestry (CEU) by the Centre d'Etude du Polymorphisme Humain (CEPH) (http://www.hapmap.org). For the third phase, the project has a set of 1397 samples and has generated 1,457,897 SNPs (http://www.sanger.ac.uk/humgen/hapmap3/). Currently, the 1000 Genomes Project is sequencing approximately 1200 anonymous volunteers from the same ethnic groups and creating a deep catalogue of human genetic variation (currently in total approximately 17 millions SNPs) (1000 Genomes Project Consortium, 2010).

In this study, the MassArray system was used to test SNPs in transcribed regions of candidate imprinted genes. Primer sets were designed within the transcribed region to allow testing of gDNA and cDNA with the same set of primers. The allelic ratios in gDNA and cDNA were therefore directly comparable as no

potential bias was introduced by primer sets of a different potency. To exclude genomic contamination of cDNA, thorough cleaning of cDNA was set up and RT negative controls systematically added to plates (see Materials and Methods section 2.5.2. for more details). The parental genotypes were determined for all SNPs tested.

The aim of the work covered in this chapter was to use the MassArray system (Sequenom, Inc.) to screen the human term placenta for imprinted gene expression using a candidate gene approach.

3.2. Detection of allelic imbalances in placental samples.

Quantitative genotyping using the MassArray system (Sequenom, Inc.) was used to test 144 genes for ASE in at least 23 family-trios. Each trio consisted of placental genomic DNA (gDNA) for identification and phasing of alleles, placental cDNA for expression levels and both parental gDNAs for determination of parent-of-origin. Six imprinted controls, seven biallelically expressed genes, eight orthologues of mouse imprinted genes, 100 orthologues of mouse imprinted candidate genes (LUEDI *et al.* 2005), and 26 human imprinted candidate genes (SEOIGHE *et al.* 2006) were assayed (See Appendix 1 Table 1 for list of genes and SNPs). The SNPs were assayed in duplicates or triplicates for four of the control imprinted genes (*DLK1*, *PEG3*, *PEG10* and *IGF2*) (error bars shown on the different graphs) but were not duplicated for all other SNPs tested to reduce costs in order to be able to screen more genes. To be informative for the detection of allele-specific expression (ASE), a SNP had to be heterozygous in the placental gDNA and expressed in the RNA. For 124 genes (86%), the cDNA amplification was successful and at least two placentas were heterozygous in gDNA. A *t*-test, followed by false discovery rate (FDR)-moderation, was used to test the null hypothesis that there was no allelic imbalance between the ratios of alleles in gDNA and in cDNA (see section 2.7.7. in Methods). The MassARRAY Typer v3.0.1 software automatically calculated the peak areas representing the quantitative 'presence' of each allele (a and b) in

the well for both gDNA and cDNA. The relative quantitation of alleles was then normalised by dividing the peak area value for one allele by the sum of the peak areas for both alleles (A=a/a+b and B=b/a+b) (see Methods). Bar plots of normalized allelic values were produced for all informative samples (Figure 23).

3.3. Controls.

Six known human imprinted genes were tested on the Sequenom platform: *DLK1*, *IGF2*, *PEG3*, *PEG10*, *ATP10A* and *PHLDA2*. For five of these controls (*DLK1*, *IGF2*, *PEG3*, *PEG10* and *PHLDA2)*, imprinted expression was confirmed and the expected parental allele was expressed. These five imprinted controls were the most statistically significant (FDR bound *p*-value) of all SNPs tested for ASE (Table 8, see next page).

The SNP rs2066707 used to target *ATP10A* was not polymorphic in our population of European ancestry and therefore could not be used to test *ATP10A* for imprinting. This was surprising as the reported frequencies of the C and G alleles were 0.342 and 0.658 respectively in the HapMap-CEU population (http://www.ncbi.nlm.nih.gov/projects/SNP/ and http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap28).
As would be expected, the average difference of expression between the two alleles in the cDNA of heterozygous individuals was greatest for imprinted genes (Table 8).

Table 8: SNPs for imprinted genes tested on the Sequenom platform.
The *p*-value was adjusted for multiple testing (FDR- bound). SR-ratio is the ratio of genotyping success rate of cDNA on gDNA. If the SR-ratio is too small, the gene is probably not expressed (see Methods section 2.7.7). The mode of ASE summarises the pattern of ASE based on the quantitative allelic expression data obtained. The fourth column lists the mean difference in expression between the paternal and maternal alleles.

Gene	SNP_ID	FDR *p*-value	Difference	SR ratio	Mode of ASE
DLK1	rs1802710	3.62E-23	88.6	0.89	Imprinting
PEG3	rs1860565	2.74E-22	98.1	1.00	Imprinting
IGF2	rs680	2.32E-16	94.6	1.04	Imprinting
PEG10	rs13073	4.19E-08	98.4	0.84	Imprinting
PHLDA2	rs13390	4.19E-08	98.1	1.13	Imprinting

In addition to numerical values analysis, a graph of normalised peak areas was produced for each gene tested to visually compare allelic ratios in gDNA and cDNA. Three examples are shown below: *DLK1*, *IGF2,* and *PEG10*. Standard errors bars are shown when the sample was tested in duplicate or triplicate.

3.3.1. DLK1.

DLK1 is paternally expressed and encodes a transmembrane protein. It is located in a cluster with *GTL2* which is maternally expressed on hChr 14q32 (KOBAYASHI *et al.* 2000; MIYOSHI *et al.* 2000; SCHMIDT *et al.* 2000; TAKADA *et al.* 2000; WYLIE *et al.* 2000). *DLK1* is expressed in the fetal liver, the adrenal gland and cortex, the fetal lung, the pituitary and the placenta (BioGPS atlas at http://biogps.gnf.org/).

Twenty-seven samples were informative for rs1802710 in *DLK1* of the 59 that were tested (Figure 18). For each trio, the placental gDNA is labelled by the family ID number followed by C (Child) and the cDNA is labelled by the same ID number followed by P (Placenta).

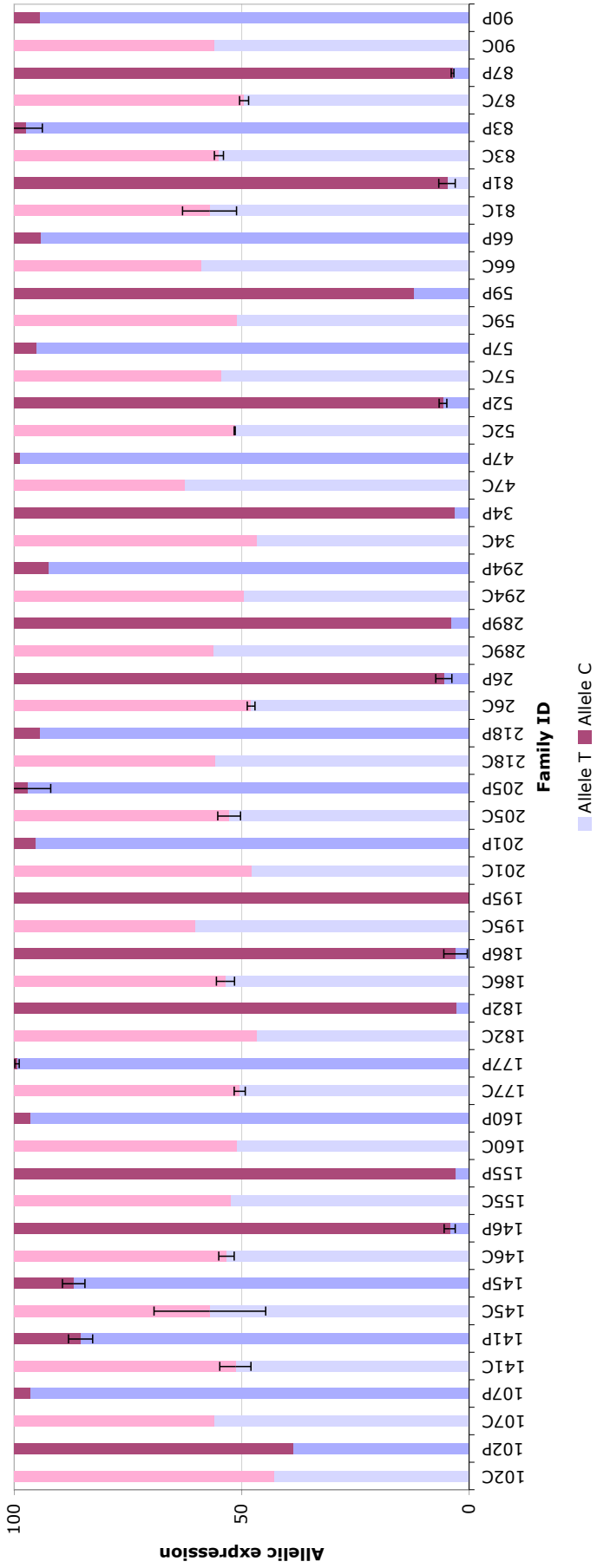

Figure 1819: Allelic expression in informative samples for rs1802710 in *DLK1*.
The normalised allelic values in placental gDNA and cDNA are plotted. First, the allelic values in gDNA (Family ID + C) (light pink and blue) and second, the allelic cDNA values (Family ID + P) (darker pink and blue) are shown next to each other for comparison.

In cDNA, either the T allele or the C allele of rs1802710 was predominantly expressed. This suggests that it is not the sequence of the allele that matters but the parent-of-origin of the allele. Using the genotypes of the parents, the paternal allele was confirmed to be the expressed allele in cDNA of all informative placental samples for *DLK1*. Twenty-six out of 27 samples exhibited monoallelic expression of the paternal allele in human term placenta.

One sample exhibited a significant expression of the usually silent maternal allele in cDNA (family ID 102 first family on the left in Figure 18). For this sample the maternal T allele was quantified at 38.7% and the paternal C allele at 61.3%. As no exon-intron boundary was crossed when the primers were designed, genomic contamination was a possible theoretical explanation for this biallelic result. However, the RT-negative control sample did not amplify, making this possibility unlikely. In this sample *DLK1* is biallelically expressed. However, the experiment was not repeated immediately and could not be repeated later.

3.3.2. IGF2.

IGF2 encodes the Insulin-like growth factor 2 at hChr 11p15.5. It is imprinted and paternally expressed (OHLSSON *et al.* 1993).
For *IGF2*, 16 samples were informative for the rs680 polymorphism. There was typical alternation of the expressed allele according to the parent-of-origin. As expected, it was the paternal allele that was monoallelically expressed in the cDNA (Figure 19).

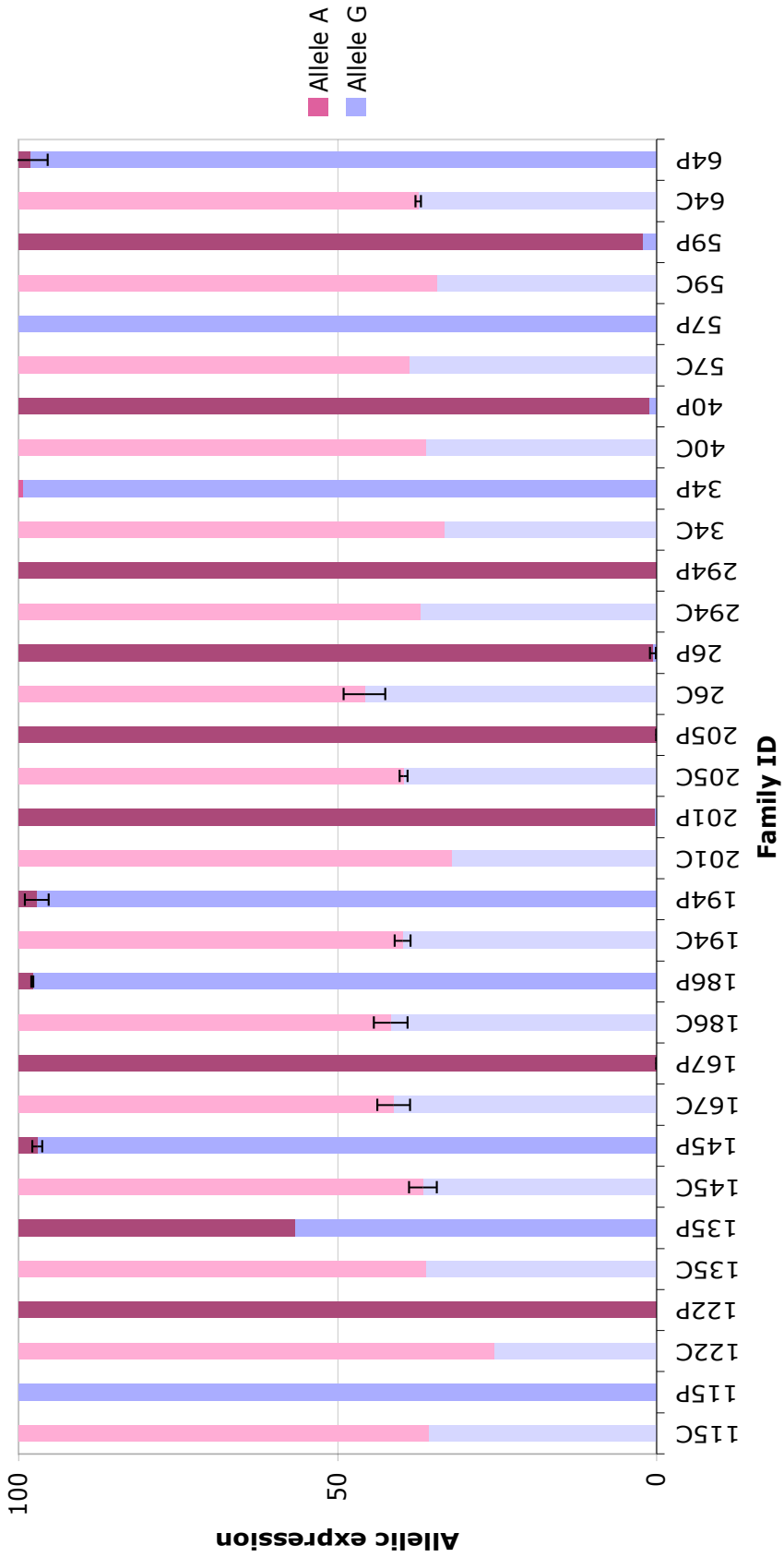

Figure 19: Allelic ratios in informative samples for rs680 in *IGF2*. Bar chart designed as in figure 18.

As for *DLK1*, biallelic expression was found in the placental cDNA of one sample (family ID 135). The normalised expression values obtained were 57 % and 43% for the G and A alleles, respectively. As both parents were heterozygous, the parent-of-origin of each allele could not be determined.

3.3.3. PEG10.

PEG10 (paternally expressed 10) is located on hChr 7q21 (ONO *et al.* 2001). *PEG10* was robustly imprinted in the eight samples that were heterozygous. The paternal allele was monoallelically expressed in all 8 cDNA samples as expected (Figure 20).

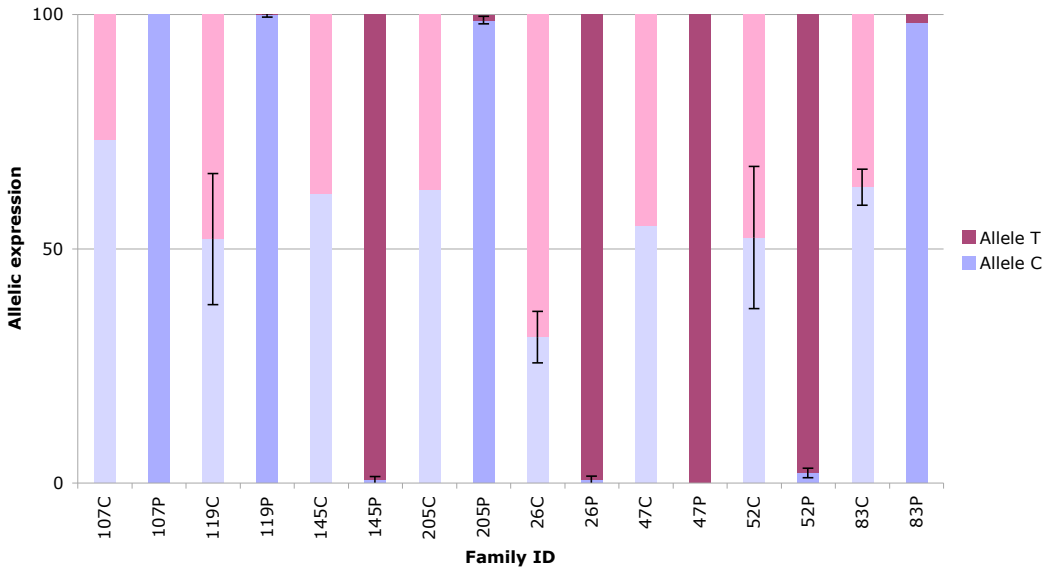

Figure 20: Allelic ratios in informative samples for rs13073 in *PEG10*.
Bar chart designed as in Figure 23.Summary of Sequenom results for imprinted genes.

The known imprinted genes, *DLK1*, *IGF2*, *PEG3*, *PHLDA2*, and *PEG10* were all confirmed as being imprinted in the human term placenta using the Sequenom assay. *PHLDA2* was maternally expressed while *DLK1, IGF2, PEG3*, and *PEG10* were all paternally expressed. Interestingly, the silenced allele was nevertheless detectable in the cDNA, while for *IGF2* and *DLK1*, one sample for each gene exhibited biallelic expression. The mean quantification of both alleles of the control imprinted genes is plotted in Figure 21 alongside the averaged

quantification of alleles for *GUSB*, a housekeeping gene tested as a biallelic control.

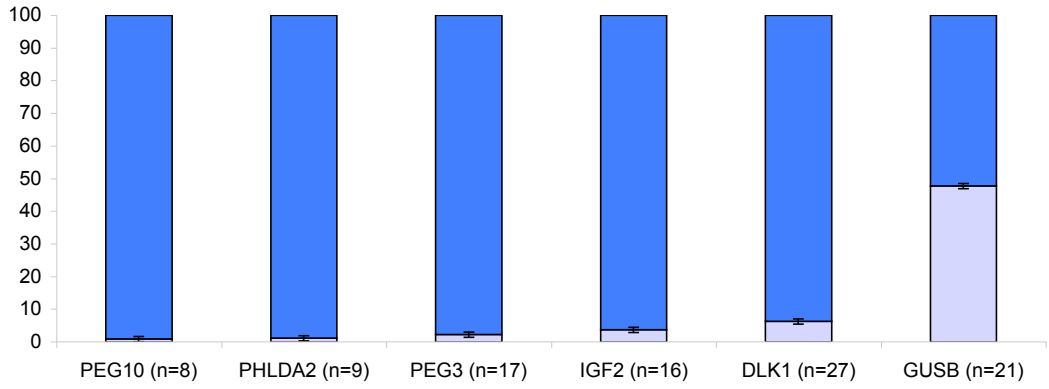

Figure 21: Lack of complete repression of the silenced allele for imprinted genes tested. Mean quantification of both alleles in cDNA across all samples tested for each of the five imprinted genes tested and for the biallelic control *GUSB*. In dark blue, the averaged values for the expressed allele and in light blue, the averaged values for the 'silent' allele. The results for the housekeeping gene *GUSB* are shown for comparison, both of its alleles are expressed equally.

3.4. Candidates tested on the Sequenom platform.

3.4.1. Screening results.

Informative SNPs were chosen in 131 candidate genes that were selected according to the criteria summarised in Figure 10 of section 2.6.2 of the Materials and Methods chapter. The placental cDNA and gDNA were tested alongside the parental gDNAs. The alleles were quantified and values obtained were normalised by dividing the peak area for each allele by the sum of the peaks for both alleles. The results were filtered according to the success rate ratio (see section 2.7.7. in the Materials and Methods chapter). The analysis suggested differential allelic expression for 6/131 SNPs (4.6%). The results for these six candidates are listed in Table 9.

Table 9: SNPs and candidate genes statistically significant for ASE.

The *p*-value was corrected for multiple testing (FDR-bound). The average difference between alleles was calculated. Genes with SR-ratio below the selected threshold were filtered out. The pattern of expression suggested by inspection of bar charts is listed in the last column.

Definition of the different effects observed:

Random ASE: one allele is more expressed than the other one in a random manner (not parent-of-origin specific)

Random monoallelic: only one allele is expressed but in a random manner

Preferential ASE: one of the alleles is always more expressed

Gene	SNP-ID	FDR *p*-values	Difference (%)	SR in gDNA	SR in cDNA	SR-ratio	Pattern of ASE
DISC1	rs821616	0.0220	15.29	95.65	86.96	90.91	Random ASE
RASGRF1	rs2230518	0.0220	75.74	86.96	82.61	95.00	Random Monoallelic
C9orf93	rs1539172	0.0398	30.03	100.00	78.26	78.26	Preferential ASE
TF	rs8649	0.0412	56.94	100.00	78.26	78.26	Random ASE
ACSS2	rs4911163	0.0412	21.93	91.30	78.26	85.71	Preferential ASE
WSCD1	rs3744725	0.0412	36.49	100.00	95.65	95.65	Random ASE

The results obtained for the candidates with an acceptable SR-ratio are detailed further in Table 10 to allow interpretation of statistical significance.

Table 10: Detailed results for ASE candidates.
SNPs, genes, chromosomal location, number of informative samples, number of failed cDNA samples, number of samples exhibiting biased expression in cDNA samples, details of pure monoallelic expression detected in cDNA samples, and suggested pattern of differential expression. For some cDNA samples, monoallelic expression was detected but the parent-of-origin could not be determined (results listed in column labelled Undet).

Gene	SNP	Chr location	Inf trios	Failed cDNA	Expr Bias	Pure monoallelic expression			Pattern of differential expression
						Pat	Mat	Undet	
DISC1	rs821616	1q42.1	11		11				Random ASE
RASGRF1	rs2230518	15q24	8	2	1	4	1		Random Monoallelic
C9orf93	rs1539172	9p22.3	12	1	11				Preferential
TF	rs8649	3q22.1	13	2	6		2	3	Random ASE
ACSS2	rs4911163	20q11.2	9	1	8				Preferential
WSCD1 /KIAA0523	rs3744725	17p13.2	11	1	10		1		Random ASE

3.4.2. Preferential expression detected in two candidate genes.

3.4.2.1. C9orf93.

For *C9orf93,* which is located on hChr 9p22.3 and which encodes an uncharacterized protein, one allele (G in rs1539172) was always more highly expressed than its alternative allele (see figures 22 and 23 for bar charts). This pattern of expression was consistent with preferential expression of the G allele. This mode of differential allelic expression was also highlighted by the statistical analysis.

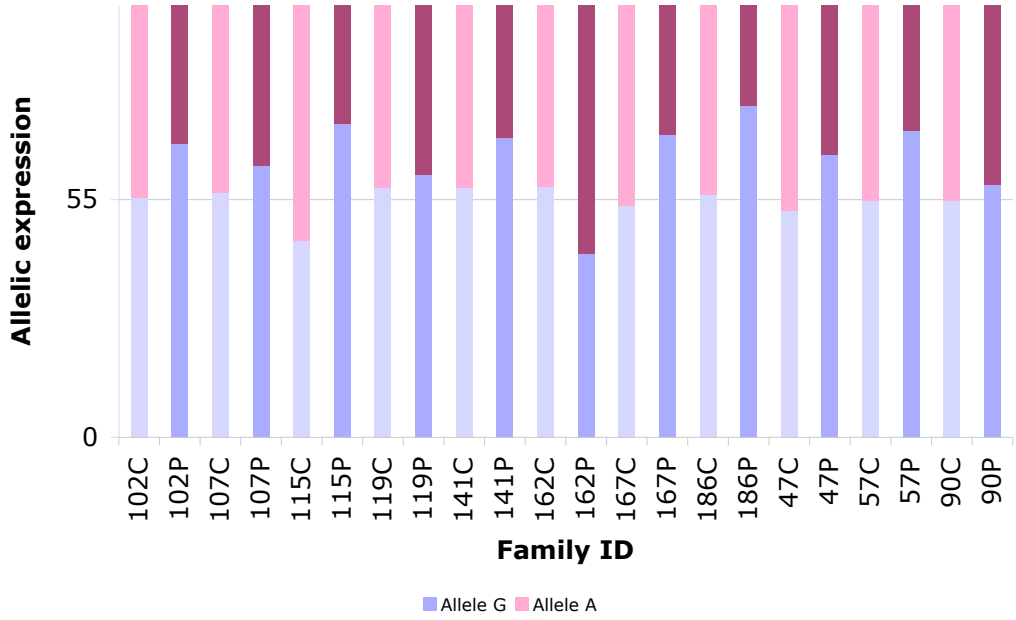

Figure 22: rs1539172-*C9orf93*.

gDNA and cDNA bar plots of the 11 informative samples. The pattern of allelic imbalance is suggestive of preferential allelic expression with the exception of Family 162. It is the allele 'G' that is preferentially expressed.

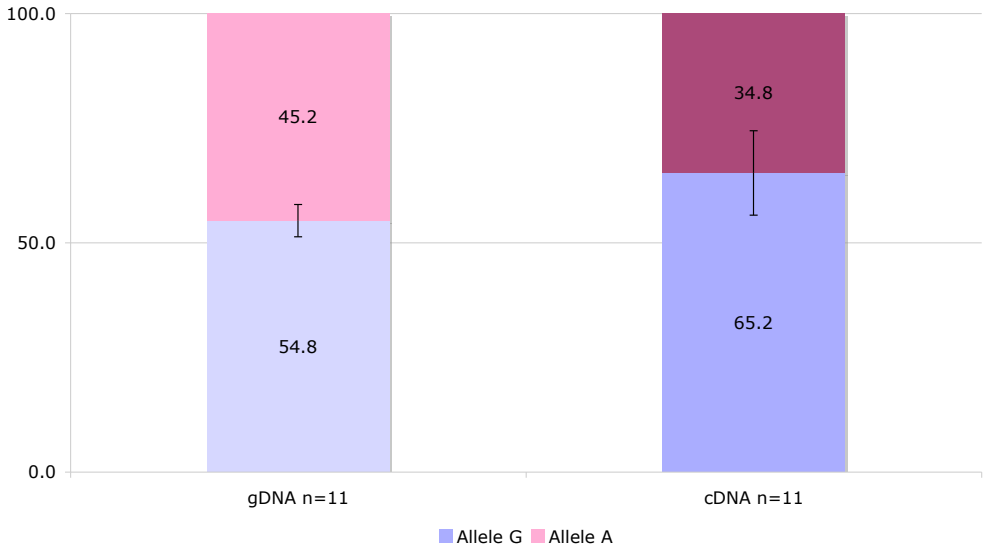

Figure 23: Preferential allelic expression of *C9orf93*.

Averaged allelic ratios for heterozygous gDNA and cDNA were plotted. The higher G/A ratio in cDNA shows preferential G allele expression (t-test *p*-value= 0.0022)

We note that there is a slight expression bias between the two alleles in the gDNA. The efficiency of the two primers is probably not exactly the same. So, the amplification step introduces an experimental bias. The strategy of using the same set of primers for both the gDNA and the cDNA has hence proved useful in order to be able to compare directly the gDNA and the cDNA as the bias is the same in both.

3.4.2.2. ACSS2.

Using rs4911163 as readout, *ACSS2* exhibited a statistically significant (two-tailed *t*-test, p=0.0075) preferential mode of ASE (Figure 24).

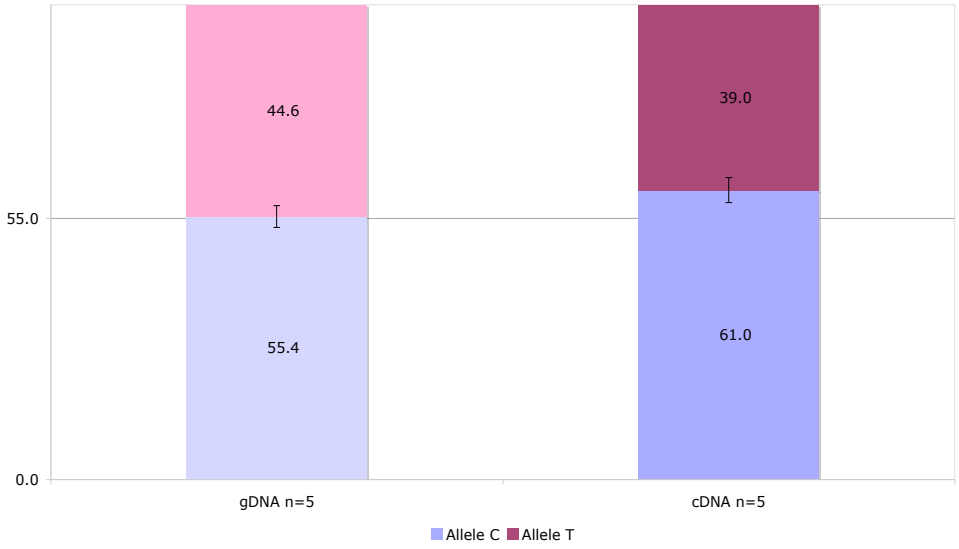

Figure 24: Preferential allelic expression of *ACSS2*.
Averaged allelic ratios for heterozygous gDNA and cDNA were plotted. The higher C/T ratio in cDNA shows preferential C allele expression (t-test *p*-value=0.0075).

ACSS2 is located on hChr 20 q11.22 and codes for a cytosolic enzyme that catalyzes acetate activation in lipid synthesis pathway. It has no known function in relation to the placenta.

3.4.3. RASGRF1.

3.4.3.1. Random monoallelic expression of RASGRF1.

RASGRF1, which is located on hChr 15q24, was found to have the most allelic difference (76%) as well as a high SR ratio (95%). The mode of expression of *RASGRF1* detected in the Sequenom experiment was compatible with incomplete random monoallelic expression (no allelic preference) (Figure 25). Monoallelic expression was detected for five out of eight informative samples (Families 103, 160, 255, 301 and 306).

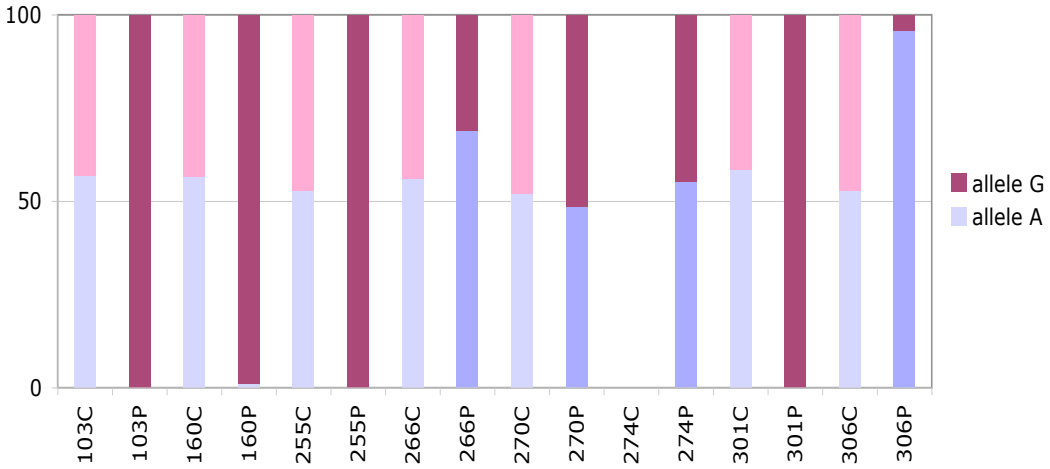

Figure 25: rs2230518-*RASGRF1.*

Bar plot designed as in figure 25. Random monoallelic expression of *RASGRF1:* monoallelic expression is detected for five out of eight informative samples (Families 103, 160, 255, 301 and 306). Paternal exclusive expression was found for four samples and maternal exclusive expression for one. The genotyping reaction failed for sample 274C but as the cDNA was heterozygous the trio was not excluded from the analysis.

The paternal allele was expressed in four of these and the maternal allele in one. The mouse orthologue of this gene is known to be imprinted in the brain and is paternally expressed (PLASS *et al.* 1996), therefore it was decided to study allele expression using another, alternative confirmatory technique.

3.4.3.2. Mode of allelic expression of RASGRF1 by direct sequencing.

To further investigate *RASGRF1* allelic expression, direct sequencing was used to test placental expression using two different SNPs: rs2230518 (=rs11855231), as previously used on the Sequenom platform and rs1562008. Both SNPs were common to all known *RASGRF1* mRNA isoforms. The primers were designed within the exon for rs1562008 but included the exon-intron boundary for rs2230518, which helps to avoid interference by genomic contamination (Figure 26).

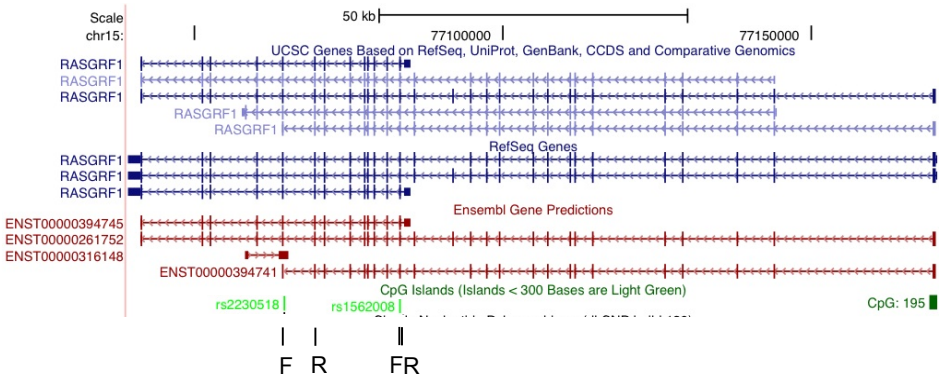

Figure 26: *RASGRF1* transcripts adapted from UCSC genome browser.
Primers used are represented as vertical bars (F and R).

Biallelic expression was found in all seven informative samples (Figure 27) with sometimes a very slight random bias between alleles.

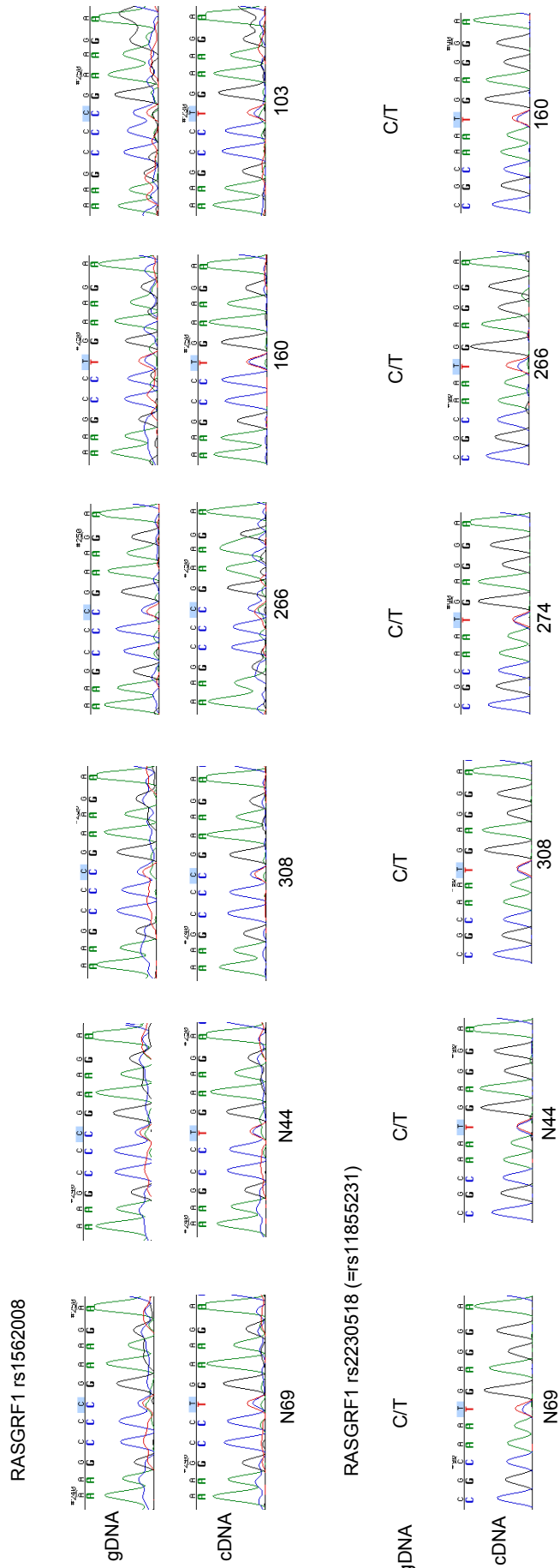

Figure 27: Sanger sequencing of two expressed polymorphisms in *RASGRF1*: rs1562008 and rs2230518.
Both alleles are expressed in the cDNA. The placental gDNA samples had been genotyped by MassSpectrometry and were not sequenced as results strongly suggested both alleles were detected.

3.5. Discussion.

This chapter has described the results obtained using the Sequenom MassArray platform. This system allows the detection and specific quantification of allelic expression at SNPs within candidate genes. Only three reactions were combined in order to quantify each allele as accurately as possible. The cost-effectiveness could have been improved further by higher order multiplexing as the system has subsequently been validated for up to 40+ SNPs per reaction (OETH *et al.* 2009). Overall, this methodology was found to be successful in testing both control imprinted genes and a first set of candidate genes.

It was of interest to note that the expression of the 'so-called' silenced allele of the imprinted control genes could be detected in some individuals in these experiments. Furthermore, *IGF2* was found to be expressed biallelically in one of 16 human term placentas (6.25%) that were tested. This was not the first time that biallelic expression of *IGF2* was reported however. Sakatani et al. studied three known imprinted genes in peripheral blood leucocytes (PBL) of 262 Japanese individuals and have found that *IGF2* was biallelic in 10% of them (SAKATANI *et al.* 2001). Roughly, the same percentages of loss of imprinting (LOI) for *IGF2* was found in control patients in a colorectal cancer study (CUI *et al.* 1998): 12% LOI of *IGF2* in normal colonic mucosa and 13% in PBL. LOI was deemed interesting as a tumour marker since it has been found in peripheral blood leucocytes for 15% of patients with colorectal cancers (CUI *et al.* 1998). However, the relevance of this finding is questionable as *IGF2* LOI is seen in healthy individuals at a similar level (SAKATANI *et al.* 2001). It will be very interesting to characterize further LOI in a much larger cohort of human samples so that the relative frequency of this phenomenon can be tested.

Our results show that *DLK1* LOI exists in human term placentas even though it only occurs rarely (1/27 placentas). Previously, *DLK1* monoallelic expression was shown to be preserved in brain tumours and lymphomas (YIN *et al.* 2004), while potential LOI of *DLK1* was described in a small number of placentas from normal and IUGR pregnancies (DIPLAS *et al.* 2009). In the latter study, *DLK1*

LOI was detected in IUGR placentas but not in five normal placentas. Its possible link to growth restriction will require validation on much larger cohorts. In our study, there seems to be very little relaxation of *PEG10* imprinting in the term placenta. This finding was corroborated by the study of Diplas et al. (DIPLAS *et al.* 2009).

On the 144 candidate genes tested, six were found to exhibit potential ASE. After detailed analysis of the results, *C9orf93* and *ACSS2* seemed to exhibit preferential ASE. However, monoallelic expression consistent with imprinting was not confirmed for any of the candidate genes tested.

The allelic expression status of *RASGRF1* has not previously been reported in humans so an orthogonal method was used to test the mode of allelic expression of *RASGRF1*. Our direct sequencing data of two SNPs demonstrate its biallelic expression (Figure 27). The average intensity obtained for *RASGRF1* in the Illumina experiment described in the next Chapter was below the cut-off value, which suggest a low level of expression in placenta. The *RASGRF1* random monoallelic expression observed in the Sequenom experiment was therefore considered to be a false positive. In mice the differential methylation of *Rasgrf1* is regulated by a repeat sequence immediately 3' of the DMD (YOON *et al.* 2002; HOLMES *et al.* 2006). As this repeat does not exist in human and it has been shown to be necessary for the imprinting of *Rasgrf1*, the biallelic expression seen in human term placenta was expected (HOLMES *et al.* 2006).

Chapter 4. Illumina® quantitative genotyping.

4.1. Introduction.

The Illumina ASE BeadArray (Illumina, Inc., USA) was used to increase screening throughput for imprinting and ASE in the term placenta. We previously used the Sequenom array to test 144 SNPs across at least 23 trios, the Illumina Array enabled us to test 1536 SNPs across the same 23 trios.

The 96-array format of the platform allows the examination of 96 samples per BeadChip. Each array is made of 1,536 different bead types. Each bead type will interrogate a different SNP and there are ~700,000 copies of a particular oligonucleotide probe covalently attached to it (see Methods section).

The ASE array is a version of the Illumina GoldenGate genotyping platform used in early experiments for gDNA and was later modified to test RNA (FAN *et al.* 2006). The array uses primer extension with allele specific primers differentially labelled with either Cy3 (green) or Cy5 (red) dye depending on which allele (allele A or B) is present at the marker position (see Methods section). This technology permits the comparison of the expression of each allele in gDNA and cDNA hybridisations. High green intensity will indicate an AA genotype while high red intensity will indicate BB genotype. If the intensity is intermediate in both channels, this will indicate an AB genotype at the marker position. We had the opportunity to design a custom made array enriched in imprinted genes and imprinted candidates (see Methods section). Several groups shared the design of the array and 357 of the 1536 SNPs tested were of particular interest to us the remainder serving as controls or additional candidates for our purposes. These SNPs were located in 214 different genes. As all SNPs on the Array were located in transcribed regions, all SNPs tested could theoretically detect imprinting since parental genotypes were also obtained. Illumina's scanning software (BeadScan) was used to quantify the fluorescence of each

bead and the software (BeadStudio)(see Methods) generates averaged intensities as the output.

The 1536 SNPs of the ASE Array (Illumina, Inc., USA) included 18 known imprinted genes, 4 housekeeping genes, 11 genes shown to be preferentially expressed in the placenta (SOOD *et al.* 2006), 10 genes predicted to be imprinted in humans (SEOIGHE *et al.* 2006), 10 orthologues of mouse imprinted genes, 35 genes that are differentially expressed according to infant weight (SOOD *et al.* 2006), 6 polycomb genes and 124 human orthologues of genes predicted to be imprinted in mouse (LUEDI *et al.* 2005). These genes were selected because they were expressed in placenta according to the expression database Unigene (http://www.ncbi.nlm.nih.gov/UniGene) (see Appendix 1 Table 2 for list of SNPs and genes). When several SNPs that met these criteria existed, SNPs with the highest MAF in our population in the single nucleotide polymorphisms database (dbSNP Build ID: 125 and 126, http://www.ncbi.nlm.nih.gov/SNP/) were chosen.

Paired gDNA (250 ng) and double-stranded cDNA (made from 250 ng total RNA, see Methods section) were hybridised to a 96-sample Sentrix Array Matrix (Illumina, Inc., USA) (FAN *et al.* 2006). Each array was scanned by Illumina BeadScan software (Illumina, Inc., USA) to produce 2 TIFF images (one for Cy3 and one for Cy5). For each placental sample, gDNA and cDNA were assayed on the same plate, and the whole plate analysis was replicated on a different day. Parental gDNA genotyping was performed on a separate plate and not replicated. The genotypes were called using Illumina's proprietary software (BeadStudio and GenCall) (Illumina, Inc.) and manually curated (see Methods section).

4.2. Imprinted controls.

First, the ASE array results were inspected for the SNPs located in known imprinted genes, some already tested by MassSpectrometry, irrespective of the cDNA hybridisation quality. Log-ratios of the hybridisation values were plotted for each informative family: placental gDNA, placental cDNA, paternal gDNA, and maternal gDNA (Figures 29, 31, 35, 37, 39 and 41). Log-ratios for heterozygous samples should approach zero on the Y-axis. For homozygous parental samples or monoallelic expression in cDNA, log-ratios will be either positive or negative on the Y-axis (see Methods). The same allele (either a or b) will always be expressed at higher levels when there is preferential expression. In the case of monoallelic expression, the allele that is more highly expressed will vary but if the parent-of-origin of the expressed allele is always the same, this pattern will be consistent with imprinting. The alternative will be considered as random monoallelic expression.

4.2.1. H19, a maternally expressed imprinted ncRNA.

H19 expresses a non-coding RNA of unknown function and is located in the imprinted cluster on hChr 11p15.5 (ZHANG and TYCKO 1992; RAINIER *et al.* 1993). The imprinting of *H19* was robust and complete in placentas from term pregnancies. The hybridisation results for rs2839702 and rs2075745 can be plotted to visually demonstrate the compact genotyping clustering and the 'pure' monoallelic expression (only two 'homozygous' clusters in cDNA) (Figure 28).

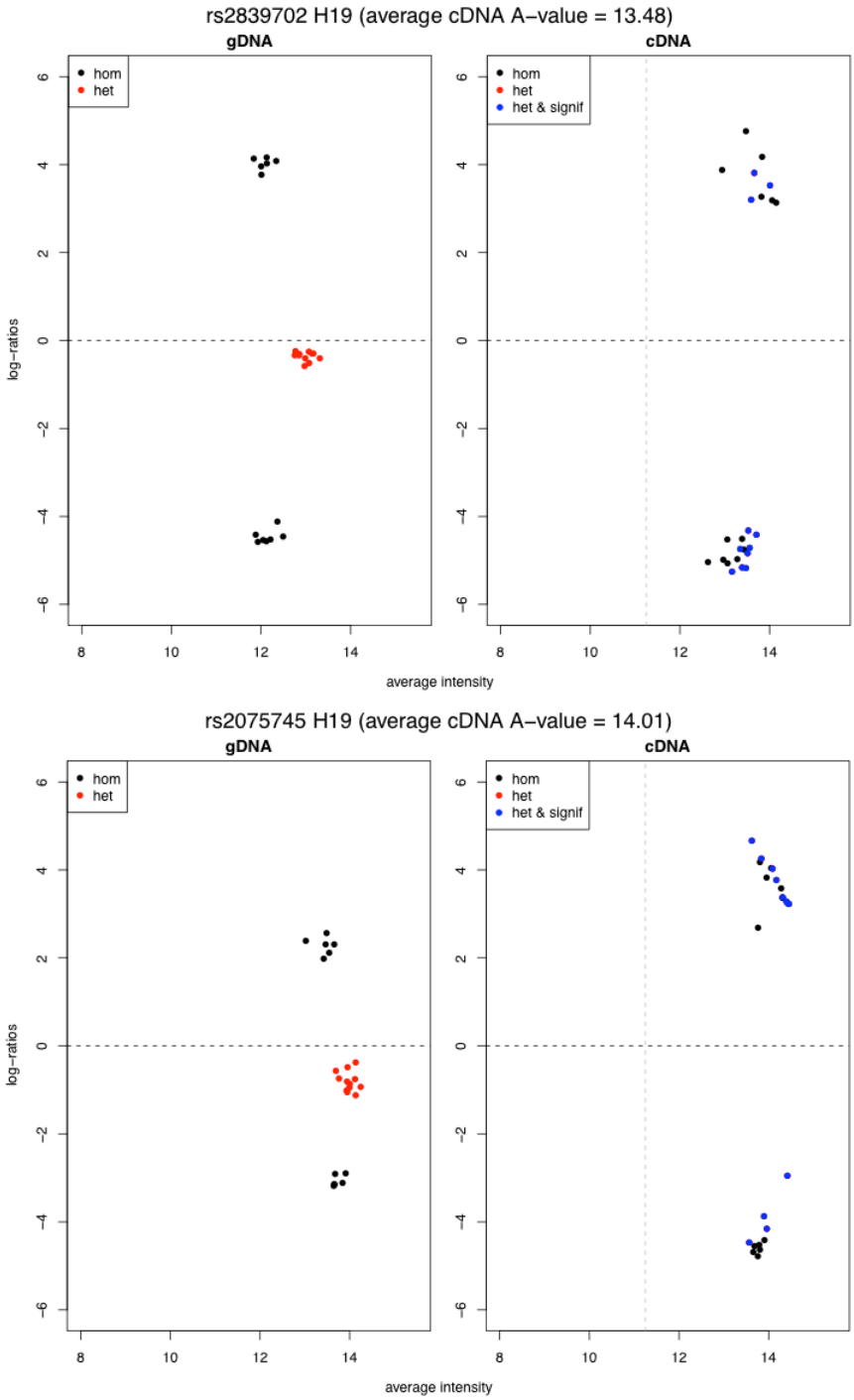

Figure 28: Scatter plots for rs2839702 and rs2075745-*H19*.

Scatter plots for the log-ratios of samples homozygous and heterozygous for rs2839702 (top panels) and rs2075745 (bottom panels). In the left panel, the homozygous uninformative gDNA samples are represented in black and the informative gDNA samples in red. In the right panel, the log-ratios of informative cDNA samples are represented in blue. The blue dots indicate that these samples are significant for ASE using the ASE statistical test (see Methods). All heterozygous samples in the case of *H19* were significant for ASE. In cDNA, the log-ratios of informative samples are consistent with monoallelic expression as the values obtained are in the same range as the values obtained for homozygotes.

For each informative SNP (i.e. heterozygous placental gDNA), allelic log-ratios were also plotted in bar charts to compare gDNA and cDNA results (Figure 29). Each family trio was grouped by a number on the X-axis and consisted of placental gDNA, placental cDNA, paternal gDNA, and maternal gDNA. In the placental gDNA, both alleles occur almost equally and the log-ratio is close to zero. The maternal allele was always the allele expressed in cDNA of parentally informative samples (i.e. at least one parent was homozygous for the marker SNP) and its signal was as high as the signal obtained for homozygous gDNA samples (Figure 29). These results were suggestive of complete imprinting of *H19* in placenta.

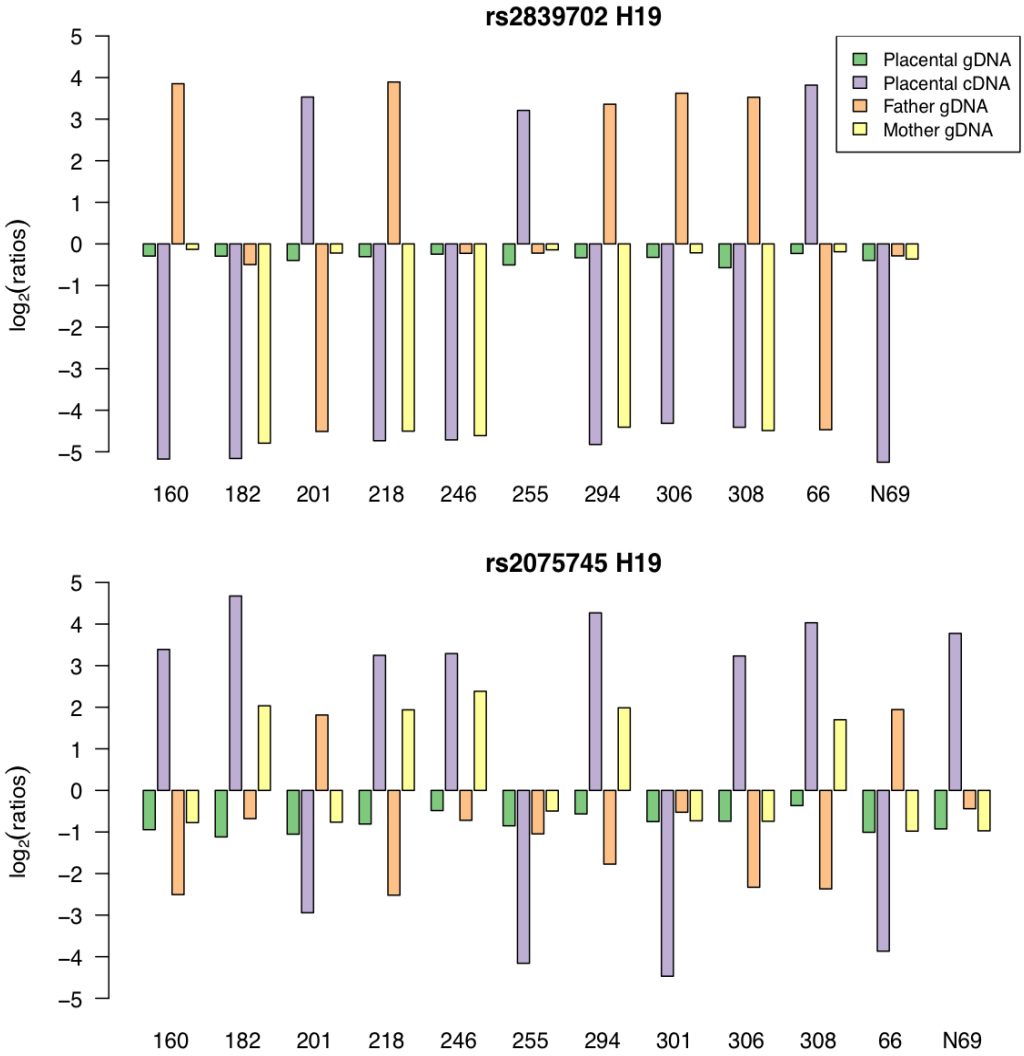

Figure 29: Informative samples for SNP rs2839702 (top) and rs2075745 (bottom) in the human *H19* imprinted gene.

Results are plotted per trio: placental gDNA in green, placental cDNA in purple, paternal gDNA in orange and maternal gDNA in yellow. In the placental cDNA, only one allele is expressed (purple bars). The sign of the log-ratio for the cDNA sample changes depending upon the allele expressed. Imprinted genes cDNA log-ratio will show a typical oscillation of signal across the y-axis because it is not the allele that is important but its parent-of-origin (POLLARD *et al.* 2008). When at least one parent is homozygous for the SNP under study, the parent-of-origin of the expressed allele can be ascertained. In the case of *H19*, the maternal allele is the one expressed as expected (i.e. for homozygous parents gDNA, the maternal gDNA allelic log-ratio has the same sign as the placental cDNA log-ratio and the paternal gDNA allelic log-ratio has the opposite sign to the placental cDNA).

4.2.2. *DLK1* is a paternally expressed, imprinted control.

DLK1 (Delta, Drosophila, homolog-like 1) (EntrezGene 8788) has been tested by Sequenom and Illumina arrays. The *DLK1*-gene was targeted on both arrays using the SNP rs1802710. For each informative placental gDNA, the corresponding cDNA exhibited monoallelic expression as expected (KOBAYASHI *et al.* 2000; SCHMIDT *et al.* 2000; TAKADA *et al.* 2000; WYLIE *et al.* 2000). The scatter plots shown are less compact in cDNA (Figure 30) than for *H19*.

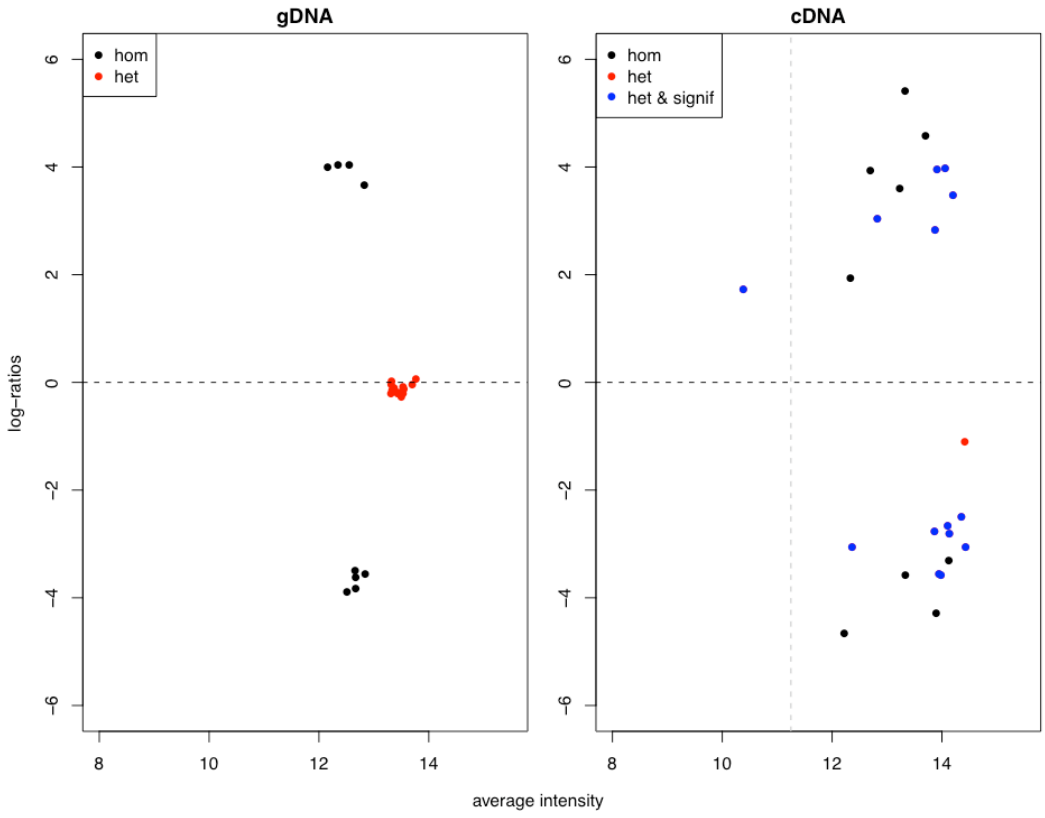

Figure 30: Scatter plots of rs1802710-*DLK1*.

Scatter plots for the log-ratios of samples homozygous and heterozygous for rs1802710. In the left panel, the gDNA informative samples are represented in red. In the right panel, the log-ratios of informative cDNA samples are represented in blue and red. The samples in blue indicate that they are significant for ASE using the ASE statistical test (see Methods). The bias observed for family 59 (unique cDNA red dot in cDNA) was not statistically significant and therefore is represented in red in the right panel.

Genotyping of the parents confirmed that the parent-of-origin of the expressed allele was the father in each case (Figure 31).

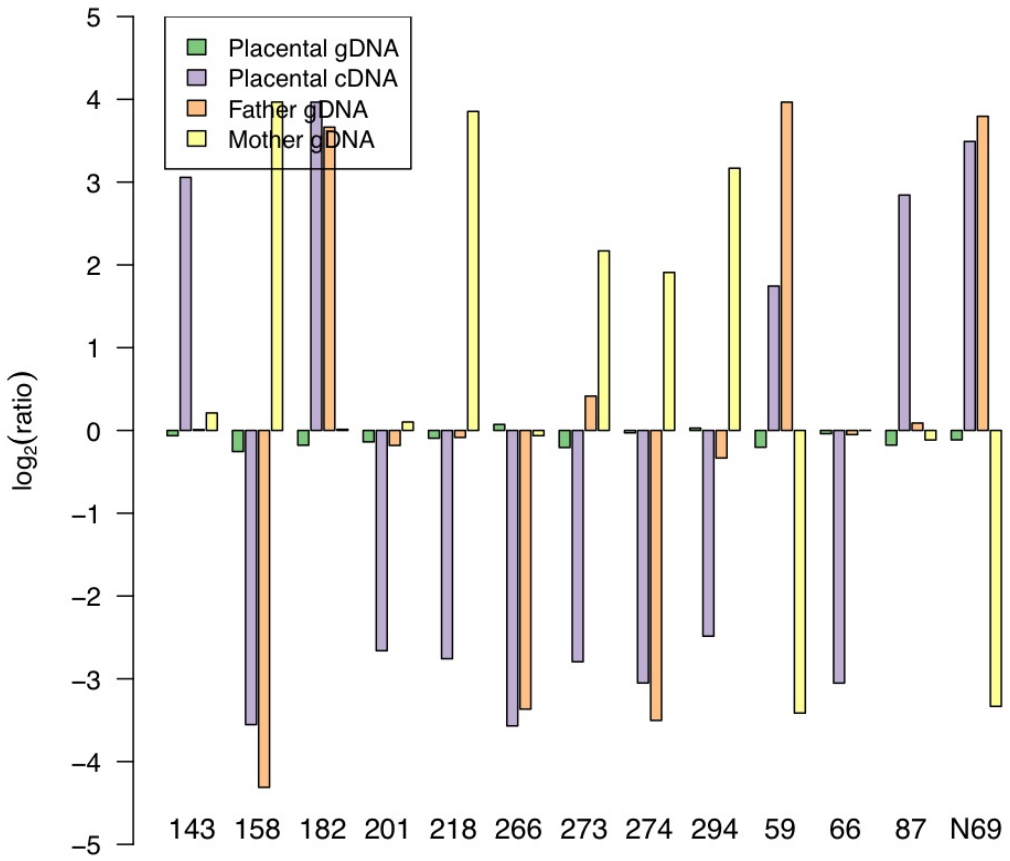

Figure 31: Bar chart of rs1802710 in *DLK1*.

There were 13 informative samples for rs1802710 in *DLK1* on the 23 tested. All informative placental gDNA samples are plotted with the corresponding cDNA and parental gDNA samples. Allelic ratios were found to be statistically significant for ASE except trio number 59. When the parental genotypes were informative, the expressed allele was confirmed to be of paternal origin (trios 158, 182, 218, 266, 273, 274, 294, 59, and N69). Bar chart designed as in Figure 29.

4.2.3. Conclusion for imprinted controls results on Illumina.

The results obtained for *H19* and *DLK1* established that the array could detect strong allelic biases and imprinting. The results for other known imprinted genes in the placenta (data not shown) were also analysed in the same way for this work. *PEG3* (MURPHY *et al.* 2001; VAN DEN VEYVER *et al.* 2001), *PEG10* (ONO *et al.* 2001), *MEST* (KOBAYASHI *et al.* 1997), *PLAGL1* (KAMIYA *et al.* 2000; ARIMA *et al.* 2001), and *IGF2AS* (OKUTSU *et al.* 2000) were all imprinted and paternally expressed while *PHLDA2* (QIAN *et al.* 1997; LEE and FEINBERG 1998) was imprinted and maternally expressed.

4.3. Comparison of platforms.

To better characterise the sensitivity and specificity of the ASE array, 38 genes were tested on both the Sequenom and Illumina platforms using the same family-trios (Figure 32).

4.3.1. Correlation of the genotypes called.

Thirty-eight genes were studied on the same 23 trios on both platforms. Firstly, just the gDNA genotyping results were analysed (paternal gDNA, maternal gDNA and placental gDNA). A total of 2568 genotypes were called on the Illumina array for gDNA, of which 2085 (81%) had a conservative rating on the Sequenom platform. Of these 2085 calls, 2074 (99.5%) were in agreement. One sample (273M) caused three discordant calls (homozygous calls on Illumina and heterozygous calls on Sequenom). On manual inspection of the ASE array clustering results, the SNP calls for this sample were identified as outliers on these three occasions. For rs11342 in the *LCP1* gene, the Sequenom platform failed to identify one of the two alleles (A) in heterozygous individuals on three occasions. The same problem was the cause of a poor correlation of results in cDNA between the two platforms for this particular SNP (see next section). A poor extension of the Sequenom primer was the likely explanation. For the last five discordant calls, different samples (66M, 40F, 289F, N44F, 273M) were identified as being heterozygous on Illumina and homozygous on Sequenom. Again a defective primer was the most likely explanation. Overall, the rate of agreement between the Illumina calls and the conservative Sequenom calls is very good.

4.3.2. Correlation of the quantitation of alleles.

Secondly, as the genotypes were in good agreement, the results of the quantitative genotyping for the 38 duplicated genes were correlated for cDNA

samples on both platforms. Pearson's correlation coefficient (r) was plotted against the cDNA intensity on the Illumina platform (Figure 32). The fall of correlation once the cDNA intensity dropped under 11.25 indicated that a cut-off value for the cDNA intensity was advisable and this was set at 11.25 for future analysis.

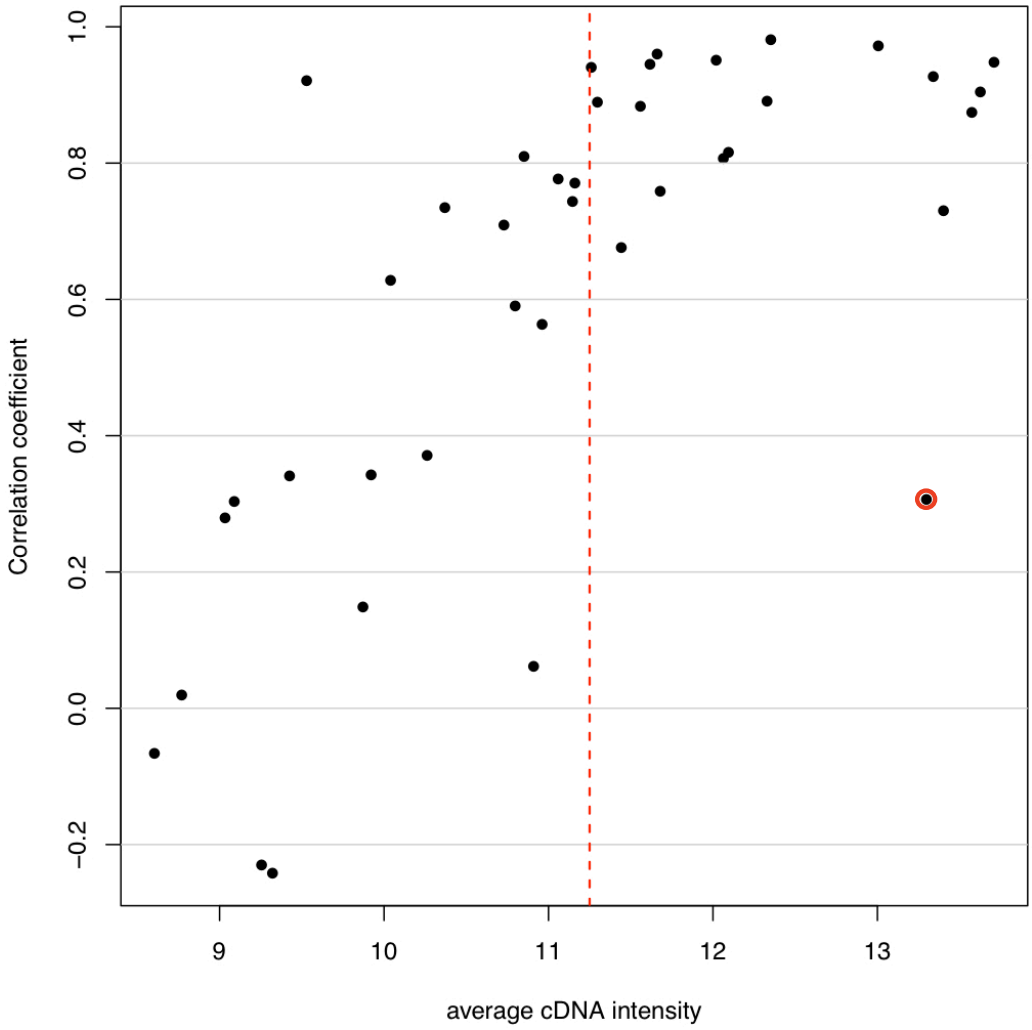

Figure 32: Correlation of Sequenom and Illumina allele quantification.
Pearson's correlation coefficient (r) was calculated for the allele-specific quantification on the two platforms and plotted against cDNA intensity (average $\log_2$ fluorescence for the 23 placentas) on Illumina. The correlation dropped when cDNA average intensity was lower than 11.25 (dashed red line). The outlier circled in red is the *LCP1* gene for which there was allelic drop-out on the Sequenom platform.

Genes that are lowly expressed in the human term placenta will yield only a small amount of transcript for the initial PCR amplification. For genes that are not expressed in placenta, there will be no input transcript: the signal detected is only noise. Hence, it was necessary to determine the lower limit of the intensity range that represented sufficient expression for reliable allelic quantification (Table 11 lists the data of Figure 32 in an algebraic way).

Table 11: Absolute Pearson's correlation coefficient |r| values for SNPs tested on both platforms and above the arbitrary intensity threshold.
The number of observations (n) is the number of heterozygous placental cDNA samples. The mean intensities obtained for the cDNA and DNA on the Illumina platform are listed in the last two columns. The mean signal obtained for cDNA was systematically lower than the mean obtained for gDNA. The gene *LCP1* (in red) had the lowest correlation coefficient despite a good intensity.

SNP	Gene	n	Absolute r	cDNA	DNA
rs754615	*CAST*	18	0.86	13.857	13.181
rs1860565	*PEG3*	12	0.98	13.796	13.471
rs1802710	*DLK1*	12	0.94	13.742	13.373
rs1050775	*NEDD9*	21	0.93	13.574	13.184
rs2304704	*SLC40A1*	23	0.94	13.512	13.097
rs11342	*LCP1*	20	0.4	13.429	13.88
rs1054124	*TGFBI*	22	0.92	13.209	12.682
rs13390	*PHLDA2*	22	0.98	12.68	12.56
rs2011951	*SLC25A39*	23	0.9	12.651	13.16
rs3803459	*RFX7*	20	0.82	12.452	13.134
rs12780	*PRDM8*	20	0.93	12.406	13.208
rs16924528	*SRP14*	20	0.83	12.385	14.428
rs9530	*GUSB*	9	0.94	12.365	13.259
rs2465811	*PTPRB*	10	0.8	12.137	12.888
rs257376	*PRKAR2B*	19	0.95	12.092	12.588
rs2425009	*MYH7B*	19	0.79	12.077	13.246
rs12214	*CTSD*	22	0.84	12.003	12.234
rs2070116	*TCF20*	22	0.72	11.925	12.905
rs2296744	*LEMD2*	23	0.85	11.795	12.988

Overall, the allelic quantification in informative cDNA samples was in good agreement between the two platforms when the average intensity across samples was greater than 11.25 (Illumina arbitrary fluorescence units).

4.4. Illumina array sensitivity for the detection of imprinting.

To assess the capacity of the Illumina ASE array to detect complete monoallelic expression, which is the most extreme form of ASE, the expression pattern of the 18 known imprinted genes on the array was analysed to develop an ASE test that could then be applied to the test SNPs (see sections 2.8.4 and 2.8.5. of Methods Chapter and Table 12). The percentage of heterozygous placentas that exhibit statistically significant ASE (Table 12) demonstrates that imprinting is reliably detected above the 11.25 average intensity threshold. The imprinting pattern is clearly less consistent (Table 12) for SNPs of imprinted genes with intensities <11.25, confirming the value of the threshold determined by the comparison of allelic expression for SNPs present on both platforms. As a consequence, the first criterion required for ASE was an average intensity across all samples > 11.25 (Illumina arbitrary fluorescence units). The second required that at least 80% of homozygotes should have adjusted *p*-values of less than 0.01 and absolute log-fold-changes greater than $\log_2(60/40) = 0.585$. This criterion was necessary to establish that the two alleles were reliably differentiated in cDNA. We note that all SNPs above 11.25 in intensity satisfy the second criterion (Table 12), while below 11.25, homozygotes were not reliably detected: the percentage of homozygotes with a p value inferior to 0.01 and absolute log-fold change (|lfc|) above 0.58 was uniformly very low.

Table 12: SNPs and imprinted control genes tested on the Illumina array.
The fore last column shows the number of heterozygous samples that were significant for ASE (p<0.01) and that had good probe hybridisation signals on the array (absolute log-fold change >0.58). The last column lists the percentage of heterozygous placentas, which exhibit statistically significant ASE. The results for *GNAS*† and *IGF2R show that these two genes are not imprinted in human term placenta.**

SNP ID	Gene	Chr	Average intensity across all samples	% of homs with p<0.01 & |lfc|>0.585	Total number of hets	Hets showing ASE: p<0.01 & |lfc|>0.585	% of hets which show ASE
rs2075745	*H19*	11	14.01	100%	12	12	100%
rs1860565	*PEG3*	19	13.63	93%	9	9	100%
rs1802710	*DLK1*	14	13.57	100%	15	14	93%
rs3730171	*GNAS*	20	13.54	96%	1	0	0%†
rs2839702	*H19*	11	13.48	100%	11	11	100%
rs998075	*IGF2R*	6	12.78	91%	13	0	0%*
rs8100247	*ZNF331*	19	12.61	91%	13	12	92%
rs9373409	*PLAGL1*	6	12.48	85%	11	9	82%
rs13390	*PHLDA2*	11	12.33	100%	2	2	100%
rs10863	*MEST*	7	12.32	95%	4	4	100%
rs12982082	*ZNF331*	19	12.27	100%	12	10	83%
rs13073	*PEG10*	7	12.09	100%	11	10	91%
rs8386	*GNAS*	20	11.97	87%	1	0	0%†
rs1055359	*PEG3*	19	11.8	100%	10	10	100%
rs1003483	*IGF2AS*	11	11.32	81%	8	7	88%
rs854541	*PPP1R9A*	7	11.2	18%	13	0	0%
rs2285185	*L3MBTL*	20	10.97	7%	10	1	10%
rs2171492	*CPA4*	7	10.57	9%	12	0	0%
rs2071970	*L3MBTL*	20	10.51	8%	11	0	0%
rs854524	*PPP1R9A*	7	10.41	21%	10	0	0%
rs8234	*KCNQ1*	11	10.18	29%	10	0	0%
rs1049846	*PLAGL1*	6	10.06	18%	13	3	23%
rs1800504	*GRB10*	7	9.98	0%	14	0	0%
rs3741208	*IGF2AS*	11	9.67	0%	10	0	0%
rs367035	*SLC22A18*	11	9.09	18%	13	0	0%
rs3816800	*ATP10A*	15	9.09	0%	15	0	0%
rs1570070	*IGF2R*	6	9.06	24%	0	0	0%
rs1800900	*GNAS*	20	8.86	7%	9	2	22%
rs2066710	*ATP10A*	15	8.8	25%	16	0	0%

Eleven imprinted control genes (tested by 15 SNPs) had a mean cDNA intensity above 11.25 units (average $\log_2$ fluorescence across the 24 placentas studied) suggesting a good expression level in term placenta. Eight genes - *H19* (Figure 29), *PEG3, DLK1* (Figure 31)*, PLAGL1, PEG10, MEST, IGF2AS* and *ZNF331* (Figure 35) - displayed a pattern characteristic of imprinting (parent-of-origin dependant monoallelic expression).

GNAS, which is maternally expressed in thyroid gland, was tested by two SNPs, rs3730171 and rs8386, which both had hybridisation intensities above 11.25. Only two placentas were heterozygous for those SNPs, and both showed biallelic *GNAS* expression (Table 12). The data suggests that *GNAS* is not imprinted in human term placenta but there were too few informative samples to ascertain this.

For *PHLDA2*, only two informative trios were available and both showed monoallelic expression as expected. The parents were informative in one trio and the expressed allele was of maternal origin. *IGF2R* was found to be biallelically expressed in 13 informative samples.

4.5. Illumina array sensitivity for the detection of ASE.

The sensitivity of the Illumina platform to detect differences in allelic expression for sufficiently expressed genes was determined by studying the precision of quantification for varying proportions of homozygous and heterozygous DNAs hybridised on the array (Figure 33 and see section 2.8.7 in Chapter 3). These 'mixture curves' showed that this platform performed well for strong ASE ratios (≥66-34).

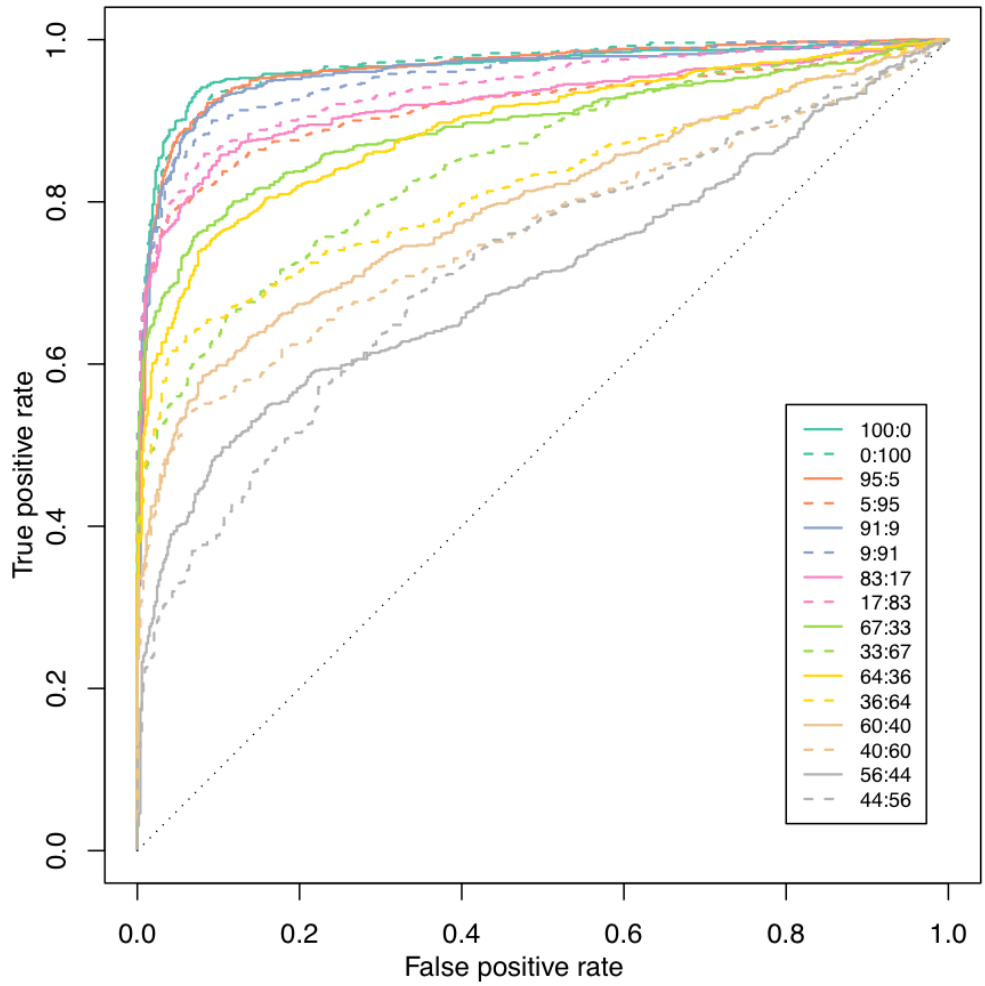

Figure 33: ROC plots for the mixture control data set.

The mean area under the 66-34/34-66 ROC curve was ≥0.81 (Table 13). The ability of the platform to detect more moderate biases (≤60-40 ratio) was weaker (area under the ROC curve ≤0.77) (Table 13, see next page).

Table 13: Area under the curve for the mixture ROC plots.

Mixture	Area under the curve	Mean area under the curve
100:0	0.9	0.905
0:100	0.91	
95:5	0.9	0.885
5:95	0.87	
91:9	0.9	0.900
9:91	0.9	
83:17	0.87	0.880
17:83	0.89	
67:33	0.84	0.825
33:67	0.81	
64:36	0.84	0.810
36:64	0.78	
60:40	0.76	0.745
40:60	0.73	
56:44	0.68	0.685
44:56	0.69	

These results show that with appropriate parameters set, the Illumina array is able to genotype gDNA samples, quantify allelic expression if the transcript is sufficiently expressed, and detect monoallelic expression and strong ASE. Therefore, an ASE test was established so that SNPs were considered to show statistically significant ASE for a particular sample/individual if they had:

- a heterozygous gDNA signal and average cDNA intensity >11.25,
- 80% of homozygous cDNA samples detectable for ASE,
- a good dynamic range across samples (absolute log fold change >0.58),
- an '*ASE statistical test*' *p*-value (false discovery bound) <0.01 for more than one heterozygous individual (see section 2.8.5).

4.6. Results of ASE statistical analysis on the Illumina ASE array.

Having established that the Illumina system could detect strong allelic expression imbalance, all the candidate and control genes were examined for evidence of ASE by the 'ASE test' described above and in the Methods (section 2.8.5). 576 out of 1536 SNPs passed the 11.25 intensity threshold indicating sufficient expression in the term placenta for reliable ASE detection (Table 14 and Appendix 1 for the full list of SNPs). Of these 576 SNPs, 497 (86%) were polymorphic in our population for at least two individuals and so were informative for the detection of ASE. A total of 261 SNPs passed the additional signal-based quality control criteria (Table 14). Using the ASE test described in the Methods section, ASE was detected in 56 out of these 261 SNPs. Of these 56 SNPs, 44 were candidate SNPs (39 genes) (Table 14 and 15) and 12 were all the imprinted controls with an average intensity >11.25.

Table 14: Number of the SNPs and genes at the various stages of the ASE array analysis.

		Genes	SNPs
A	Tested on the array	932	1536
B	Above intensity threshold (11.25)	446	576
C	As in B with at least two heterozygous samples	393	497
D	As in C with good quality probe hybridisation in homozygotes	214	261
E	As in D with at least two heterozygotes significant for ASE ($p<0.01$)	49	56
F	As in E for the candidate genes only	39 (18.2%)	44 (16.9%)

Five different types of ASE were identified manually in our results: (1) imprinted, i.e. monoallelic expression in a parent-of-origin dependent manner; (2) ASE in a parent-of-origin manner: 'partial imprinting'; (3) preferential ASE: the same allele is expressed at higher levels in each heterozygote irrespective of its parent-of-origin; (4) random monoallelic expression: one of the two alleles is randomly completely silenced; (5) random ASE, alleles are randomly expressed at higher levels in different heterozygotes without any specific parental bias (Table 15). To determine which of these patterns of allelic imbalance in expression was detected, log-ratios of informative family-trios were plotted as described in Figure 34 and the plots examined (data not shown). The patterns of allelic imbalance identified for the 56 SNPs are reported in Table 15.

Table 15: List of SNPs and corresponding genes significant for ASE (p<0.01).
All SNPs had an average intensity above 11.25 and a number (No) of heterozygous samples had a significant statistical ASE test (p<0.01).
The pattern of ASE was determined by visual examination of bar plots of the quantitative genotyping results for paired cDNA and gDNA samples.

rsID	Name	Chr	Control or candidate status	Average intensity all >11.25	No of hets	No of hets with p<0.01	Pattern of ASE
rs1802710	*DLK1*	14	impr control	13.57	15	14	imprinting
rs8100247	*ZNF331*	19	impr control	12.61	13	12	imprinting
rs2075745	*H19*	11	impr control	14.01	12	12	imprinting
rs2839702	*H19*	11	impr control	13.48	11	11	imprinting
rs1082	*PHACTR2*	6	candidate	12.09	14	10	partial imprinting
rs12982082	*ZNF331*	19	impr control	12.27	12	10	imprinting
rs13073	*PEG10*	7	impr control	12.09	11	10	imprinting
rs1055359	*PEG3*	19	impr control	11.8	10	10	imprinting
rs9373409	*PLAGL1*	6	impr control	12.48	11	9	imprinting
rs1860565	*PEG3*	19	impr control	13.63	9	9	imprinting
rs2309428	*TJP2*	9	candidate	12.86	9	8	random ASE
rs178077	*SNAP29*	22	candidate	11.95	10	7	random ASE
rs1003483	*IGF2AS*	11	impr control	11.32	8	7	imprinting
rs8585	*UBE2V1*	20	candidate	12.36	13	6	preferential
rs1130663	*CD151*	11	candidate	12.44	18	5	random ASE
rs4664114	*FMNL2*	2	candidate	11.51	14	5	random ASE

rsID	Name	Chr	Control or candidate	Average intensity	No of hets	No of hets with p<0.01	Pattern of ASE
rs4944960	*XRRA1*	11	candidate	12.43	12	5	preferential
rs2282336	*TJP2*	9	candidate	12.32	9	5	random ASE
rs6633	*CDK2AP1*	12	candidate	11.34	8	5	random ASE
rs4614	*VPS11*	11	candidate	13.09	13	4	random ASE
rs3817672	*TFRC*	3	candidate	12.07	13	4	preferential
rs2905	*C14orf130*	14	candidate	11.33	12	4	random ASE
rs12190287	*TCF21*	6	candidate	12.88	10	4	random ASE
rs3809865	*ITGB3*	17	candidate	12.09	10	4	random ASE
rs10863	*MEST*	7	impr control	12.32	4	4	imprinting
rs915894	*NOTCH4*	6	candidate	13.44	16	3	Preferential ?
rs754615	*CAST*	5	candidate	13.71	14	3	preferential
rs838896	*SCARB1*	12	candidate	12.62	10	3	random ASE
rs5758651	*TCF20*	22	candidate	12.4	10	3	random ASE
rs11699879	*NCOA3*	20	candidate	12.68	9	3	random ASE
rs838891	*SCARB1*	12	candidate	11.85	9	3	random ASE
rs2425009	*MYH7B*	20	candidate	11.68	9	3	random ASE
rs9749449	*ZNF211*	19	candidate	11.71	6	3	random ASE
rs4797	*SQSTM1*	5	candidate	14.48	18	2	preferential
rs1128933	*MAN2C1*	15	candidate	12.71	16	2	preferential
rs10277	*SQSTM1*	5	candidate	11.78	16	2	preferential

rsID	Name	Chr	Control or candidate	Average intensity all >11.25	No of hets	No of hets with <0.01	Pattern of ASE
rs2249057	*NM_006031*	21	candidate	13.31	13	2	preferential ?
rs7226091	*MGC16597*	17	candidate	11.83	13	2	preferential ?
rs1043618	*HSPA1A*	6	candidate	13.6	12	2	random ASE
rs17085249	*ELL2*	5	candidate	12.98	12	2	random ASE
rs2255255	*CRNKL1*	20	candidate	12.86	12	2	random ASE
rs2013162	*IRF6*	1	candidate	12.64	12	2	preferential ?
rs11121567	*PGD*	1	candidate	12.03	12	2	preferential ?
rs3780473	*ACO1*	9	candidate	11.67	12	2	random ASE
rs4669	*TGFBI*	5	candidate	13.27	11	2	random ASE
rs7242	*SERPINE1*	7	candidate	13.01	11	2	random ASE
rs2788478	*FLJ10300*	7	candidate	13.04	10	2	random ASE
rs2271108	*DOCK5*	8	candidate	12.85	10	2	random ASE
rs552282	*PPFIA1*	11	candidate	11.58	10	2	preferential ?
rs1044116	*NOTCH3*	19	candidate	12.21	9	2	random ASE
rs7204628	*MGC24665*	16	candidate	11.84	9	2	random ASE
rs844	*FCGR2B*	1	candidate	13.05	8	2	preferential ?
rs11156878	*KIAA0391*	14	candidate	12.15	5	2	random ASE
rs12780	*PRDM8*	4	candidate	12.02	5	2	random ASE
rs5919	*ITGB3*	17	candidate	11.55	4	2	random ASE
rs13390	*PHLDA2*	11	impr.	12.33	2	2	imprinting

Imprinting is the ASE pattern that is easiest to detect by visual inspection or statistical analysis. Random monoallelic expression was not detected in the pool of SNPs analysed. Preferential ASE was detected and was frequent. The remaining SNPs were categorised as exhibiting random ASE. This last category is the one most likely to contain false positives. Also as the number of heterozygous samples significant for ASE was very small for more than half of the SNPs, the results for these SNPs must be treated with caution. However, this is a very conservative analysis. Indeed, for *SQSTM1* for example, there are only two samples out of 18 that were statistically significant for ASE for rs4797 (data not shown) and 2/16 for rs10277 (see Figure 41), but preferential expression of one of the two alleles was clearly evident on inspection of the bar plots (Figure 41 page 162). The same SNPs were also previously shown to exhibit preferential expression in LCLs (YANG *et al.* 2010).

The analysis of ASE results for the imprinted controls was detailed in sections 4.2., 4.3., and 4.4. The subsequent sections present analysis of the most interesting ASE results for candidate genes.

4.6.1. *ZNF331* is imprinted in human placenta.

The most promising imprinted gene candidate is the *ZNF331* gene (Table 15). It encodes a zinc finger protein on hChr19q13.41 (RefSeq NM_018555). It was studied with two different SNPs: rs8100247, located in exon 1 (part of the 5'UTR of the gene) and rs12982082 in exon 2 (also in the 5'UTR) (Figure 34). The pattern of allelic expression was typical of imprinting as the expressed allele was always inherited from the same parent-of-origin, the mother (Figure 35). These results strongly suggest that the *ZNF331* isoforms targeted by the SNPs used are imprinted and maternally expressed in human placenta.

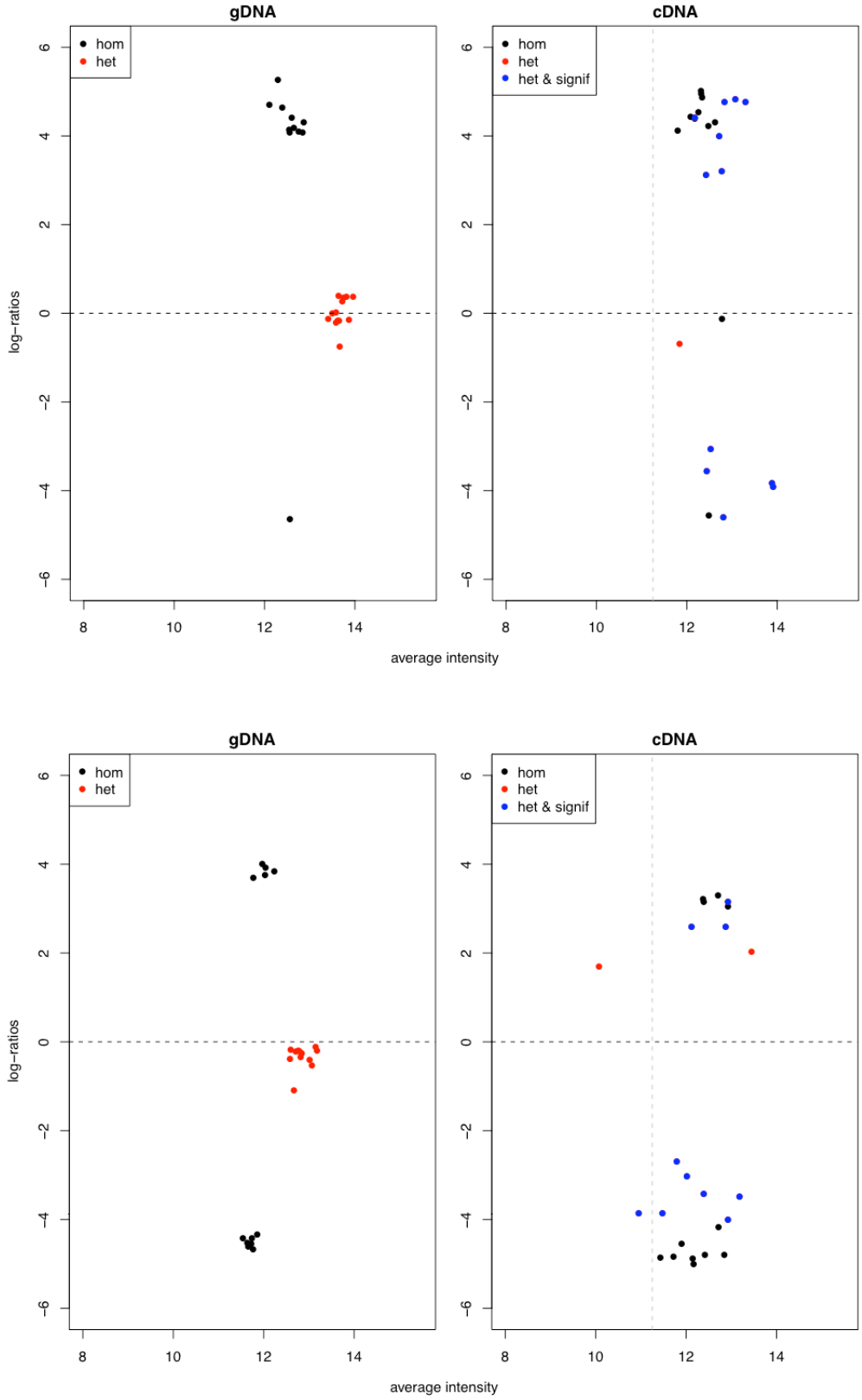

Figure 34: Scatter plots for *ZNF331*-SNPs.

rs8100247 (top panels) and rs12982082 (bottom panels). There is typical splitting of the log-ratios in cDNA (blue and red dots in right panels) in two clusters for the heterozygous individuals (red dots in left panels). Scatter plot designed as in Figure 28.

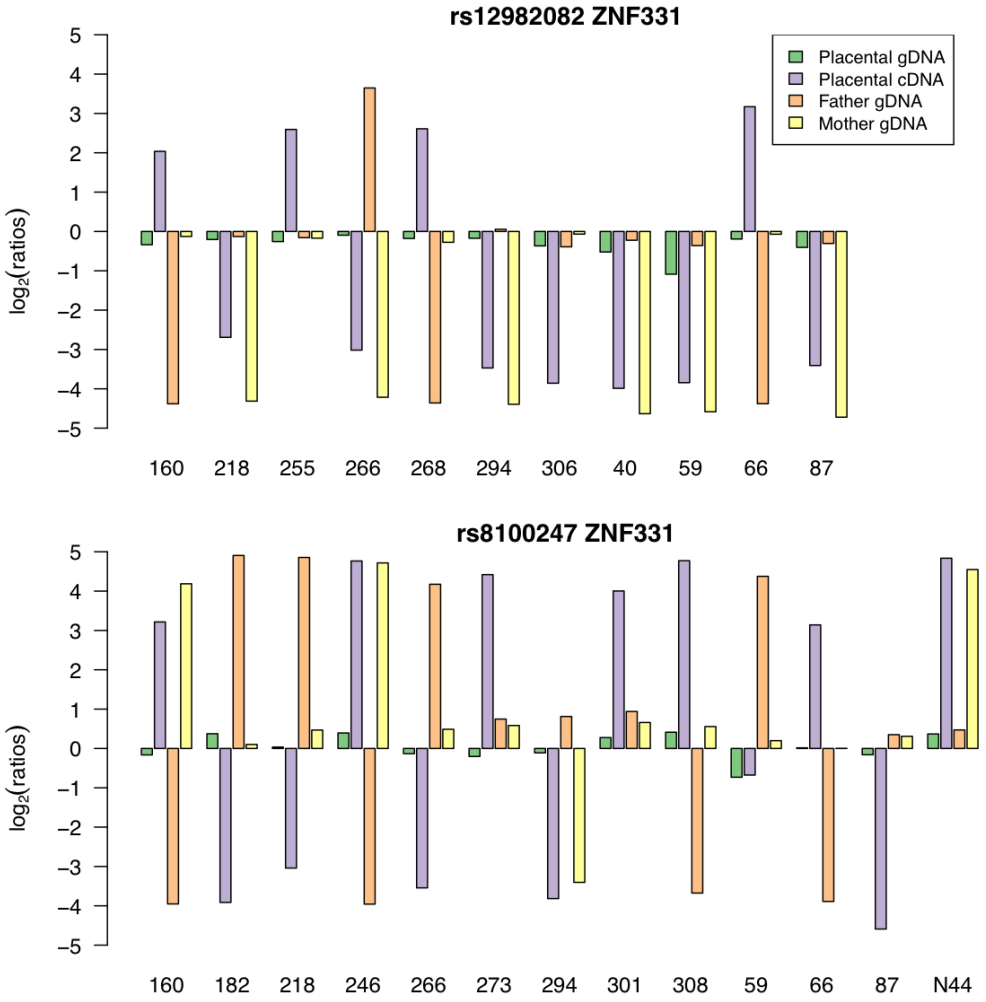

Figure 35: ASE array analysis: Bar charts for rs12982082 and rs8100247 in *ZNF331*.
Monoallelic maternal expression of *ZNF331* was detected for two SNPs rs12982082 and rs8100247. The log_2(ratios) of heterozygous placental gDNA samples are represented along corresponding cDNA and parental log_2(ratios). There is typical 'imprinted gene' oscillation of the signal across the Y-axis. The expressed allele is inherited from the mother. Bar chart designed as in Figure 29.

4.6.2. *PHACTR2* is partially imprinted in placenta.

Another promising candidate for imprinted expression was *PHACTR2* (phosphatase and actin regulator 2 gene) (Table 15). The *PHACTR2* gene is located on hChr6q24.2. It is 114,245 bp distant from *PLAGL1,* a known imprinted gene (also known as *ZAC*). The ASE array tested allelic expression of

the rs1082 SNP, which is located in the 3'UTR of the gene, and of rs2073214 located in the fifth exon.

For rs1082, 10 of 14 informative samples exhibited significant ASE dependent on the parent-of-origin (Figures 36 and 37). The expressed allele along the Y-axis and parental genotypes suggested that it was always the maternal allele that was more expressed (Figure 37). However, the log-ratios in cDNA were always smaller than the ones obtained for homozygous individuals. These smaller cDNA log-ratios suggest partial imprinting.

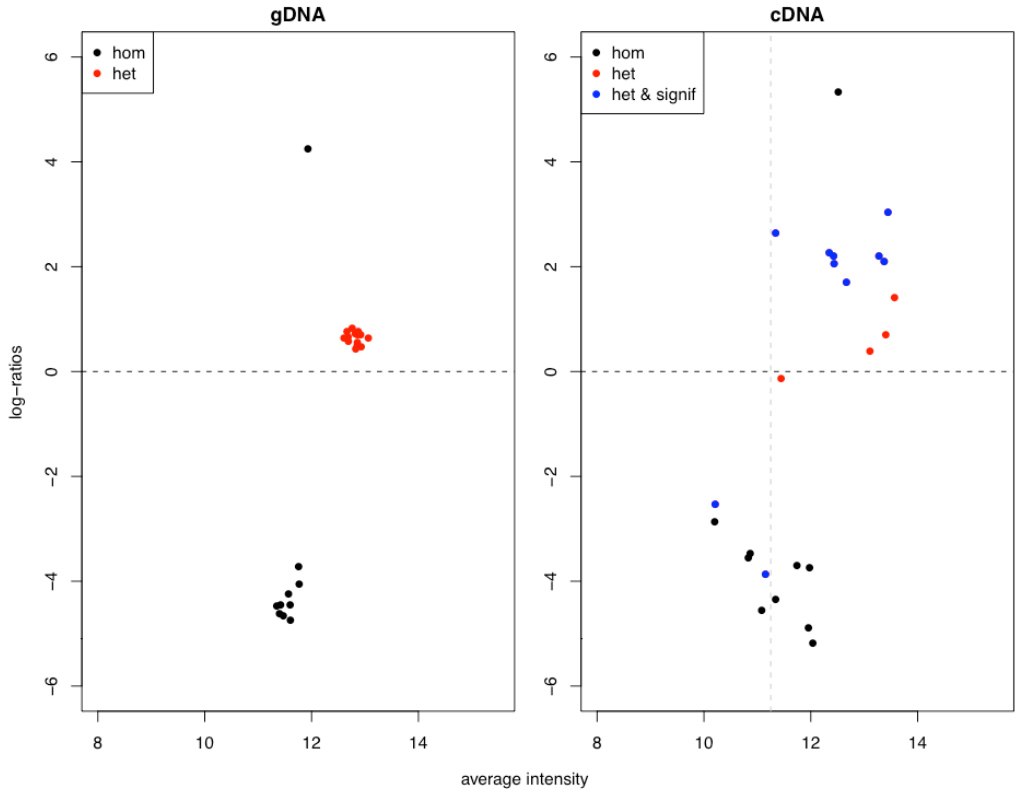

Figure 36: ASE array analysis: Bar plot for rs1082-*PHACTR2*.
There are several samples that are significant for ASE. However, the cDNA ratios are smaller than the ones obtained for homozygous samples, which suggests partial imprinting. Scatter plot designed as in Figure 33.

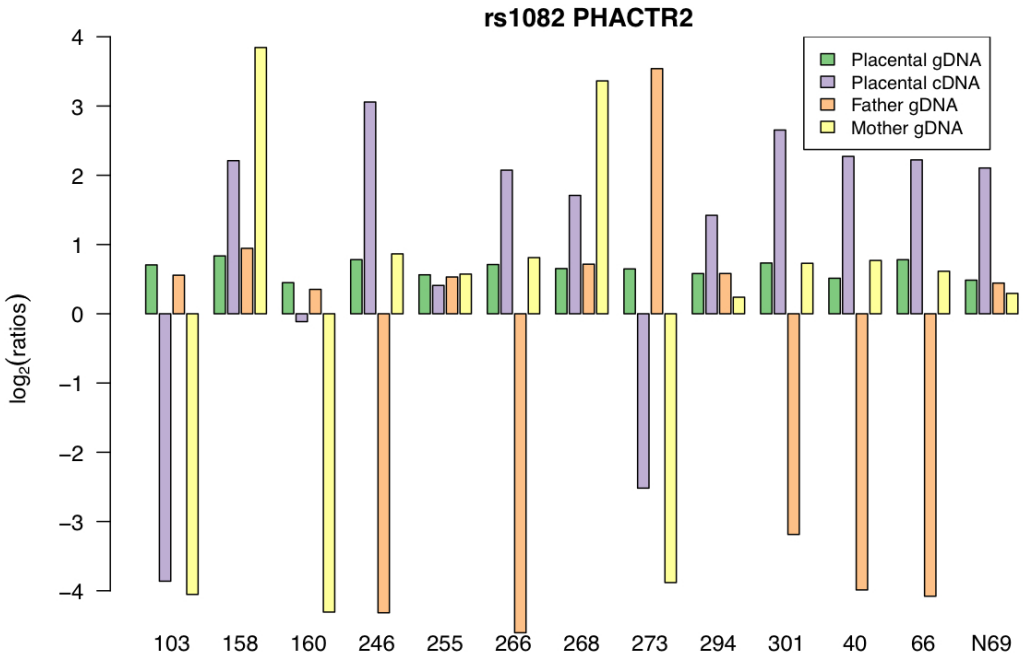

Figure 37: ASE array analysis: Bar chart for rs1082-*PHACTR2*.
The ASE pattern of rs1082-*PHACTR2* is consistent with partial imprinting (maternal expression bias). Bar chart designed as in Figure 29.

The analysis of results for the other marker SNP rs2073214 in *PHACTR2* demonstrates how difficult the detection of less obvious ASE patterns can be (Figures 38 and 39, see next page). There were fewer informative samples (six heterozygous placentas) and the bias was smaller than for rs1082. When a bias was present (Figure 39) and the parents were informative, the maternal allele was expressed at a higher level (samples 143, 246, and 273). However there is only one statistically significant sample (273) (Figure 38), and this SNP was not listed among the most promising candidates, which all required two statistically significant samples for ASE (Table 15, see page 150).

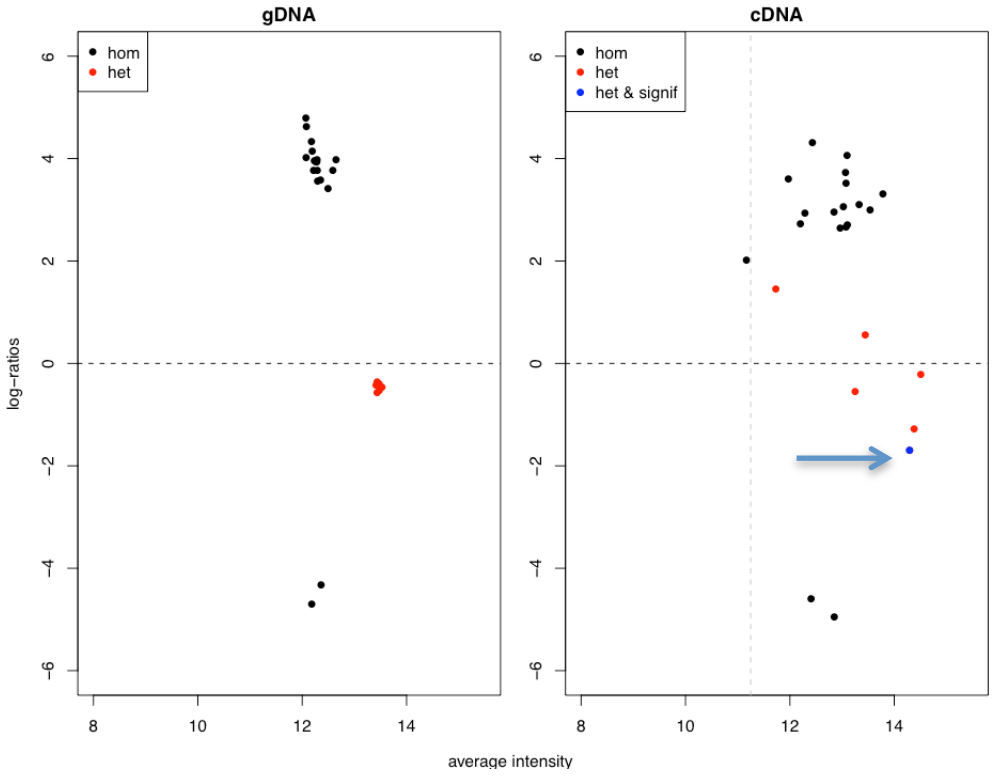

Figure 38: ASE array analysis, scatter plot for rs2073214-*PHACTR2*.
One sample exhibits statistically significant ASE (blue dot pointed by the blue arrow). The spread of the log-ratios for the other informative samples is large. Scatter plot designed as in Figure 28.

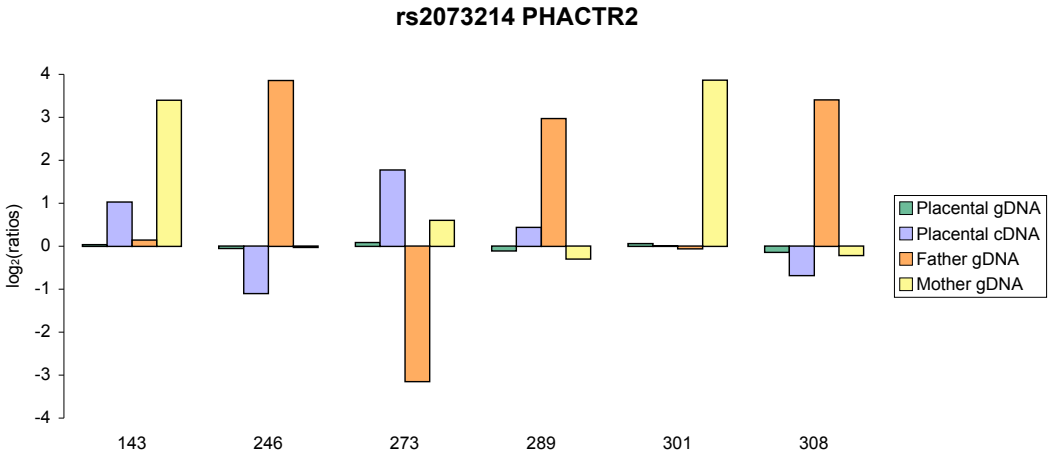

Figure 39: ASE array analysis, bar chart for rs2073214-*PHACTR2*.
There is a biased expression towards the maternal allele but the allelic ratio is smaller than for homozygous samples. Bar chart as in Figure 29.

Two SNPs per gene were chosen on the array when possible and SNPs not in linkage disequilibrium were preferred. The two SNPs tested for *PHACTR2* are

64,188 bp distant from each other. The SNP rs2073214 is located in an exon that is subject to alternative splicing and this SNP exhibited a lesser degree of partial imprinting than rs1082 (Figures 37 and 39). This suggests that differences in allelic biases between the different isoforms of *PHACTR2* could exist.

4.6.3. The spectrum of silencing for imprinted genes.

To examine the strength of allelic silencing observed in our data for all the known imprinted genes and the most significant imprinted candidate genes, raw allelic values were averaged over all cDNAs from informative individuals and plotted (Figure 40). The difference of expression between the two alleles of the control imprinted genes varies from a 23-fold difference (*PEG3* -rs1860565) to a 6.4-fold difference (*DLK1*- rs1802710). For the confirmed imprinted gene *ZNF331*, the difference is 5-fold for rs12982082 and 11-fold for rs8100247, and 2.6-fold for the partially imprinted gene *PHACTR2* (Figure 40). These results show that the repression of the silenced allele is not complete for all imprinted genes and that there is a continuum from ‘complete imprinting’ to ‘partial imprinting’.

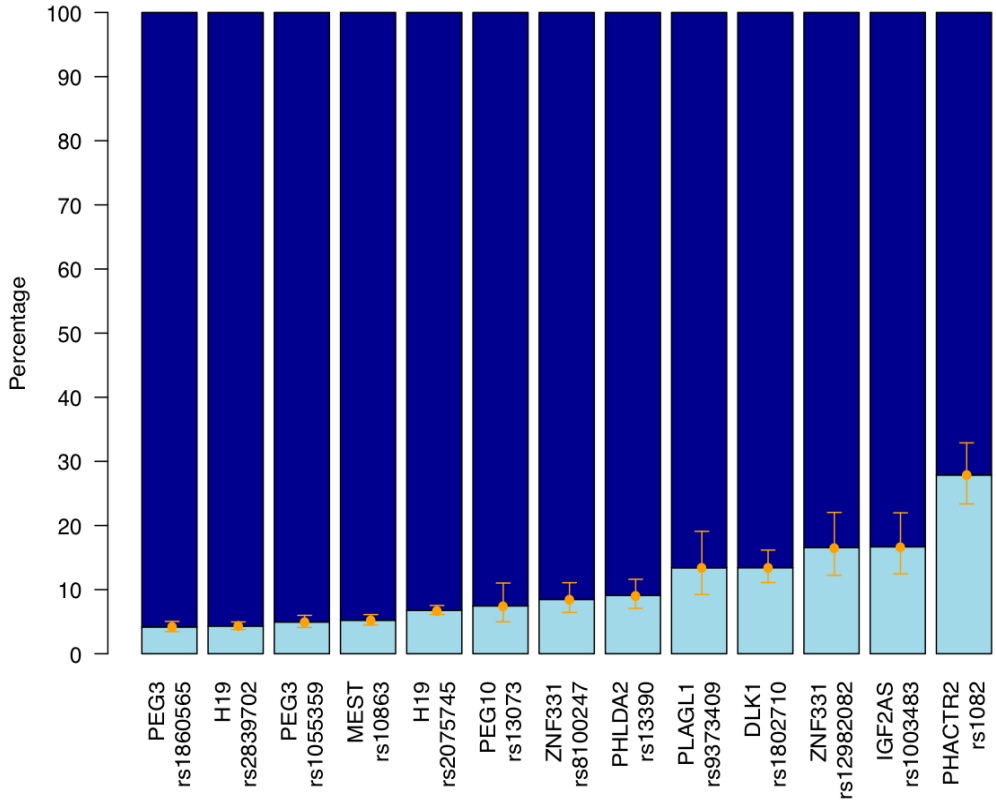

Figure 40: Lack of complete repression of the silenced allele for imprinted genes.
Average quantification of the expressed allele (dark blue) and of the silenced allele (light blue) in all informative samples for all expressed imprinted controls (*PEG3, H19, MEST, PEG10, PHLDA2*, *PLAGL1, DLK1 and IGF2AS*), for *ZNF331* and for *PHACTR2*. The means (orange dots) and standard errors (orange bars) were calculated across all heterozygotes in logit scale to avoid intervals less than zero and greater than one (see Methods section page).

While our results could suggest that it is likely that most or all 'completely imprinted' genes have already been found in the placenta (see discussion), our *PHACTR2* study indicates the possibility that partially imprinted genes could have been labelled as 'biallelic' and that several other similarly imprinted genes could still be found and characterised.

4.6.4. Candidates statistically significant for other ASE types.

4.6.4.1. Candidate genes showing preferential expression of one allele.

From the 56 SNPs statistically significant for ASE, ten SNPs were located in eight of the imprinted genes (*DLK1 H19*, *PEG10*, *PEG3*, *PLAGL1*, *IGF2AS*, *MEST*, *PHLDA2*), two in *ZNF331* and one SNP was located in *PHACTR2*. Of the remaining 43 SNPs, six SNPs (five genes) showed an allelic preferential pattern when visually examined: *UBE2V1*, *XRRA1*, *CAST*, *SQSTM1* and *MAN2C1* (Table 15 on page 150 and Figure 41, see next page). For eight further SNPs the pattern was also possibly compatible with preferential expression. In these cases, either the number of informative samples or the biases observed were insufficient to be conclusive (SNPs dubbed 'preferential?' in Table 15, see page 150).

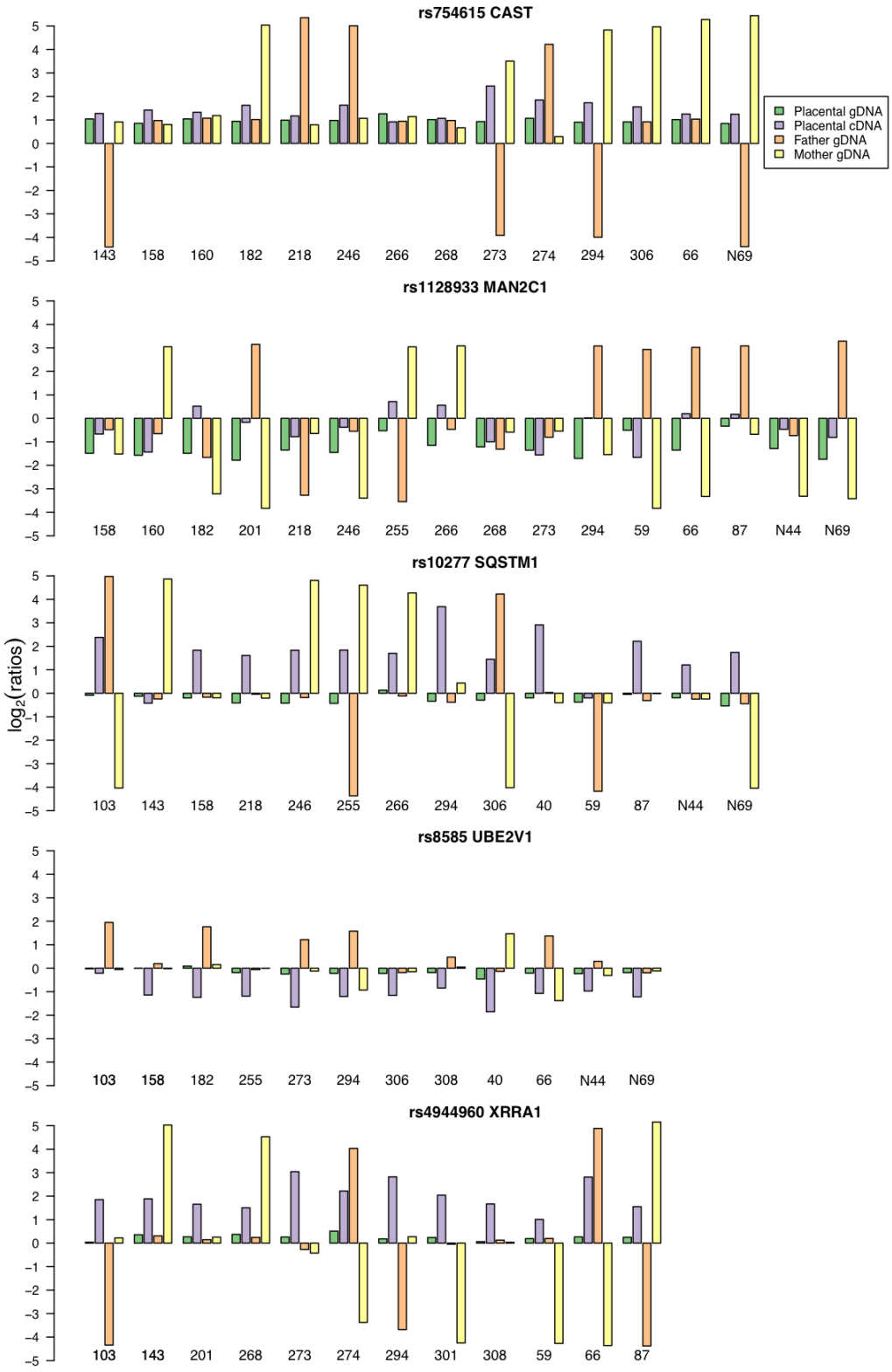

Figure 41: Preferential expression of one allele.

The ASE for SNPs in *SQSTM1*, *UBE2V1*, *XRRA1* (the last three panels) is evident while it is more subtle for SNPs in *CAST* and *MAN2C1* (the top two panels).

Preferential allelic expression is concluded when one of the two alleles is always more expressed than the other. This is clearly seen using graphical representation where the bars are always on one side of the y-axis (Figure 41). Preferential allelic expression was the most common form of allelic imbalance that we found (9 genes on the Illumina array using stringent criteria). For six of these preferential expression had been detected in lymphocytes (STRANGER *et al.* 2005). Hence, preferential allelic expression is the most common mode of ASE in placenta.

4.6.4.2. Candidates showing other forms of ASE.

Eight showed possible allelic preference (Table 15, see page 150). The other SNPs were too variable to be assigned a precise ASE pattern and their expression could correspond to random allelic bias, epistatic allelic preferential expression, bipolar ASE (CHEVERUD *et al.* 2008) (see Chapter 6, Discussion, page 199) or false positives.

4.6.5. Results for imprinted mouse genes with an unknown status in human.

For all mouse imprinted genes with an unknown imprinted status or conflicting data in human at the start of the study, we demonstrated biallelic expression in human placenta (Table 16). These results were expected for *IGF2R* and *SLC22A3* (MONK *et al.* 2006), and *PON2* (MONK *et al.* 2008).

Table 16: List of mouse imprinted genes with unknown status in human tested on the Illumina and Sequenom arrays.
The only gene for which no reports existed is *RASGRF1*. It proved to be biallelic on Illumina and by Sanger sequencing (see previous Chapter).

Gene	Chr location	Imprinting in human (Otago imprinted gene catalogue)	Results
IGF2R	6q25.3	Monoallelic in 3/8 placentas (MONK *et al.* 2006)	Biallelic in 13 placentas
SLC22A2	6q26	Polymorphic (5/18 monoallelic in placenta) (MONK *et al.* 2006)	Expression too low in placenta
SLC22A3	6q26	Imprinted in first trimester placenta (MONK *et al.* 2006)	Biallelic in 10 informative placentas
CALCR	7q21	Provisional data (monoallelic in 4/5 brains) (MONK *et al.* 2008)	Expression too low in placenta
PON2	7q21	Biallelic in 4 term placentas (MONK *et al.* 2008) and other tissues	Biallelic in 10 informative placentas
DHCR7	11p13.4	Two placental samples (biallelic) (SCHULZ *et al.* 2006)	Expression too low in placenta
AMPD3	11p15.4	One placental sample (biallelic) (SCHULZ *et al.* 2006)	Expression too low in placenta
TH	11p15.5	No reports	Expression too low in placenta
GATM	15q21	Biallelic (MONK *et al.* 2008)	Expression too low in placenta
RASGRF1	15q24	No reports	Biallelic in 18 informative placentas
USP29	19q13	No reports	Expression too low in placenta
ZIM3	19q13	No reports	Expression too low in placenta
ZNF264	19q13	No reports	Poor probe hybridisation

As *RASGRF1* was highlighted as having ASE by the Sequenom study, we studied it by Sanger sequencing and demonstrated that it is biallelically expressed in human term placenta (see previous Chapter for details).

4.7. Discussion.

The Illumina Beadarray™ platform enabled the correct identification of known imprinted genes and was reliable in the detection of strong allelic skewing (>66-33) as shown. For some genes, the allelic ratio in gDNA departed from the 50:50 ratio probably because the efficiency of both primers were not equal. Hence, the strategy to use the *same* set of primers to quantify alleles in gDNA and cDNA (i.e. RNA) was crucial.

Of 119 candidate genes for which data was obtained on Sequenom, 6 were statistically highlighted as being differentially expressed (5.0%). Of the 214 candidate genes (261 SNPs) that passed quality control tests on Illumina, 39 candidate genes (18.2%) and 44 SNPs exhibited ASE. This higher ASE detection rate on the Illumina platform could be explained by a gene selection bias, by noisier results and/or a more stringent analysis of Sequenom results.

On the Illumina platform, one candidate, *PHACTR2*, was found to be partially imprinted. The other candidates showing ASE seemed to be preferentially (13 genes) or randomly expressed (26 genes). The first category represents certainly more reliable results, the latter could be meaningful results or noise. Alternative technologies, such as direct RNA sequencing, may be needed to validate the latter category.

Considering the expected allelic expression obtained for the control imprinted genes and for known differentially expressed genes, the hardware used seems trustworthy. However, some pitfalls were evident. The Illumina platform generated higher throughput than the Sequenom platform but with the cost of lower sensitivity and increased noise. Indeed, genes expressed at a ‘medium or

low' level in placenta were not reliably detected meaning that far fewer genes could be tested than anticipated. Data from other groups confirm that this was due to the Illumina GoldenGate technology itself: hybridisation arrays are known to be less sensitive than other platforms for quantification of PCR products.

Furthermore the Illumina technology relies on hybridisation and this step potentially introduces inaccuracy in the experiment (allele-specific biases). This assay was originally designed for DNA genotyping and later adapted to RNA quantitative genotyping (FAN *et al.* 2003; FAN *et al.* 2004). The synthesis of cDNA is also a source of supplementary variability (for both platforms). However despite its noise, the GoldenGate technology was used successfully to test ASE in lymphoblastoid cell lines (LCLs) (SERRE *et al.* 2008b) and pancreatic cancer cell lines (TAN *et al.* 2007).

This array was custom made to test many candidate genes in human term placenta. Of the 1536 SNPs (targeting 932 genes) present on the Illumina array, 44 SNPs (39 genes) exhibited statistically significant ASE. Three SNPs were located in two potentially 'new' placental imprinted genes that are studied further in the next Chapter: *PHACTR2* and *ZNF331* (Chapter 5).

While on a quantitative scale, these results are slightly disappointing, on a qualitative scale, they allowed us to study allelic expression in a native tissue on an unprecedented scale.

Chapter 5. Confirmation of imprinting for *PHACTR2* and *ZNF331*.

5.1. Confirmation of the partial imprinting of *PHACTR2* at the *PLAGL1/HYMAI* locus by Sanger sequencing.

5.1.1. Introduction.

PHACTR2 was identified as a partially imprinted gene by the screen on the Illumina array. When one of the two alleles was more expressed, it was always the maternal one (Figures 37 and 39 in the previous Chapter). This gene lies 100 kb from *PLAGL1* (*ZAC*), which is a known imprinted gene. *PLAGL1* is responsible for transient neonatal diabetes mellitus (TNDM; MIM 601410), which is characterised by neonatal hyperglycemia requiring insulin treatment, intra-uterine growth restriction and neonatal failure to thrive. *PLAGL1* encodes a zinc finger transcription factor involved in insulin secretion control. The vast majority of TNDM cases are due to 6q24 defects: paternal UPD, paternal duplication or loss of maternal methylation at the locus DMR. Recently a TNDM patient with a hemizygotic deletion encompassing *PHACTR2*, *PLAGL1*, *HYMAI*, *SF3b5* and *STX11* was described (DIATLOFF-ZITO *et al.* 2007). In addition to TNDM, the patient had multiple congenital abnormalities (heart defects, bronchodysplasia, delayed bone maturation) (DIATLOFF-ZITO *et al.* 2007). Another known imprinted gene is located at 6q24: *HYMAI* (hydatidiform mole associated and imprinted), which is untranslated, of unknown function and paternally expressed (ARIMA *et al.* 2000). *HYMAI* overlaps the first exon of *PLAGL1* (Figure 42). The CpG island that also overlaps exon 1 is the imprinting control region for the locus (ARIMA *et al.* 2006).

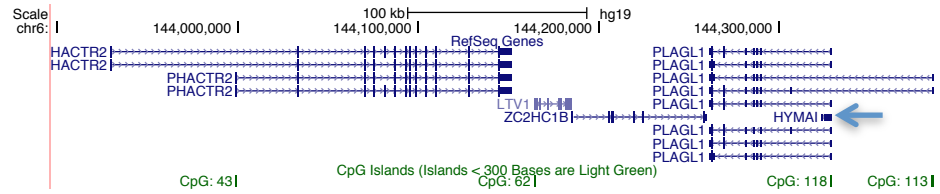

Figure 42: *PHACTR2* locus.

***PLAGL1* and *HYMAI* are overlapping. *HYMAI* is indicated by a blue arrow on the figure. The CpG 118 is the ICR for the locus. Figure generated with the UCSC website.**

Four transcript variants are listed in RefSeq Genes for *PHACTR2* (Figure 42). The marker SNP rs1082 (Figure 43) was tested on the Illumina array and was significant for ASE. The variants 1 (NM_001100164.1) and 3 (NM_014721.2) share an additional exon that contains rs2073214 that was also tested on the ASE array and for which one sample out of six informative ones was statistically significant for ASE (the maternal allele was more highly expressed) (Figure 39 in the previous Chapter).

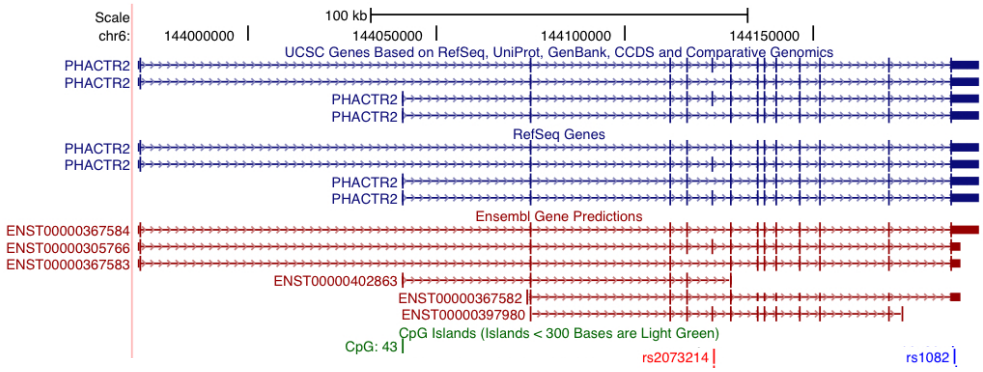

Figure 43: *PHACTR2* known transcripts.

Figure generated with UCSC genome browser. SNP rs1082 is located in the 3'UTR and rs2073214 is in the fifth exon, which is subject to alternative splicing. Both were tested on the Illumina ASE array.

5.1.2. Study of Sanger sequencing traces for *PHACTR2.*

5.1.2.1. Sanger sequencing for rs1082-PHACTR2 in term placenta.

Partial imprinting of *PHACTR2* was confirmed using Sanger sequencing on nine informative placental samples. A recurrent maternal bias was seen between gDNA and cDNA sequence traces overlapping the same *PHACTR2* 3'UTR SNP (rs1082) (Figure 44, see next page).

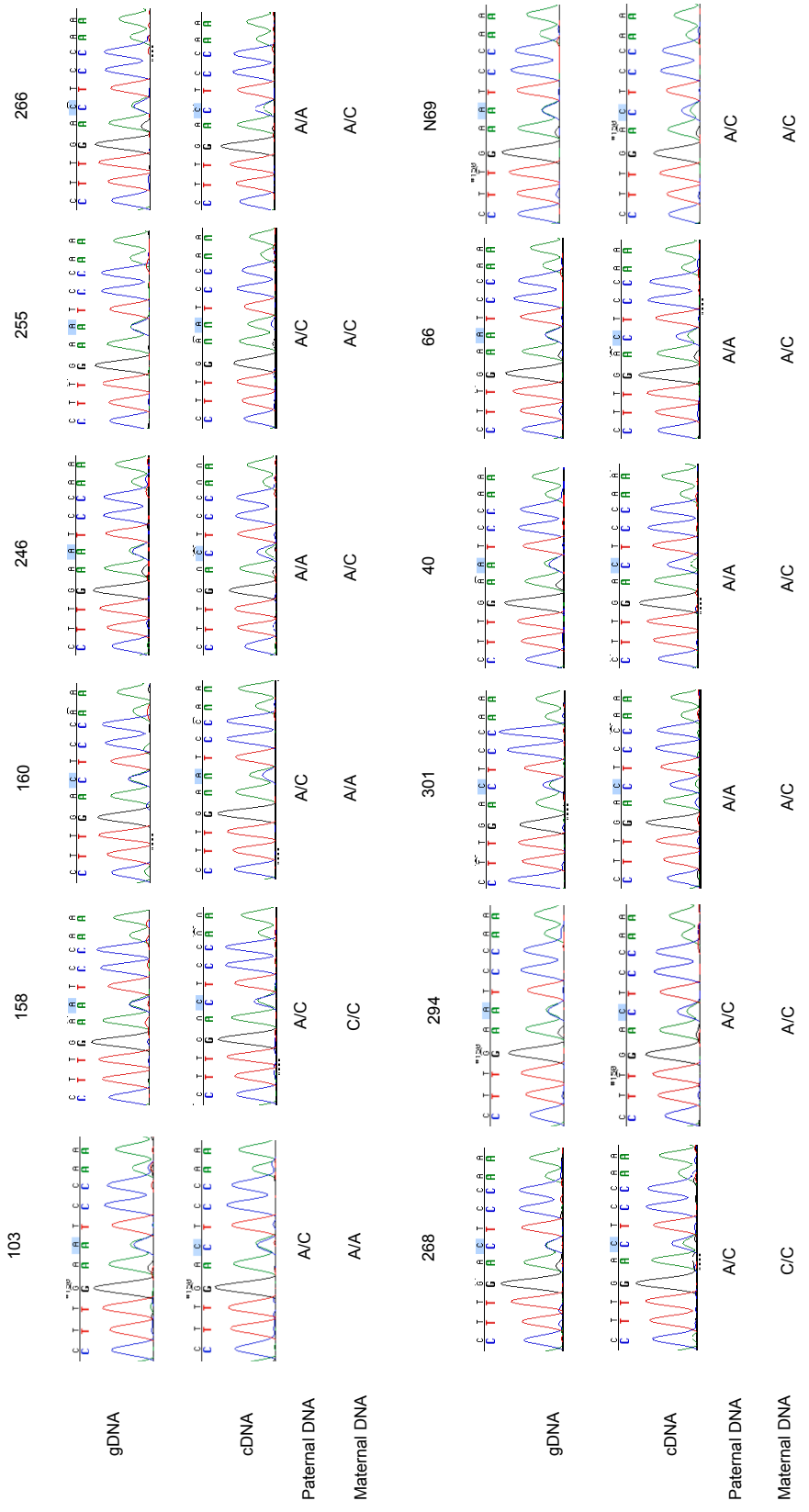

Figure 44: Sequences traces rs1082-*PHACTR2*.

Nine informative placental samples were sequenced with the same set of primers for gDNA and cDNA. For each trio, the sequence traces are grouped with the Illumina genotyping data for the parents. A recurrent maternal bias is seen between gDNA and cDNA traces.

These results confirm the partial imprinting of *PHACTR2* in human term placenta and the ability of the Illumina BeadArray™ platform to detect ASE. As imprinting could vary with the stage of the pregnancy, we tested *PHACTR2* imprinting in the first trimester trophoblast.

5.1.2.2. Sanger sequencing in first trimester placentas.

5.1.2.2.1. *Analysis of rs1082 in first trimester placentas.*

The same polymorphism (rs1082) was studied in a set of trophoblast samples. Four informative individuals were studied. The maternal bias, even though smaller, was present in three out of four individuals in the first trimester (Figure 45).

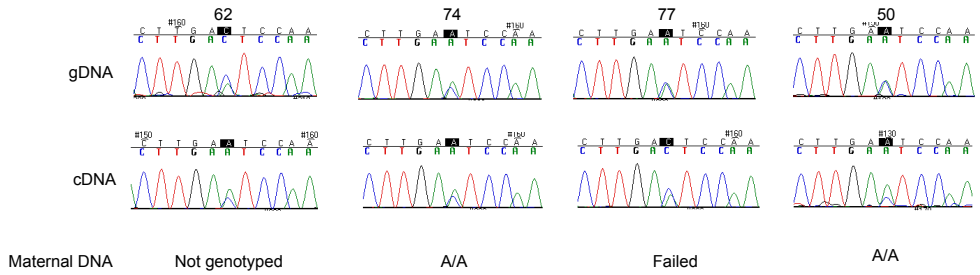

Figure 45: Sequences traces of four informative trophoblast samples for rs1082-*PHACTR2*.

5.1.2.2.2. *Study of rs2073214 in first trimester trophoblasts.*

The rs2073214 SNP was also studied on the Illumina array. The results obtained were suggestive of partial imprinting (Figures 39). This SNP was sequenced in informative first trimester trophoblasts. There is overexpression of the maternal allele (Figure 46).

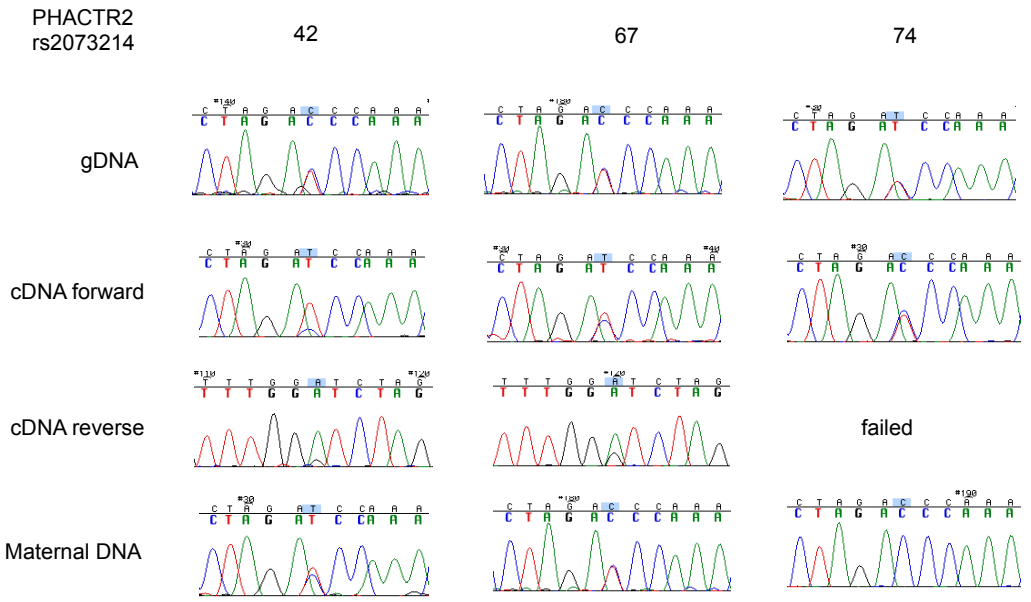

Figure 46: Sequence traces of rs2073214-*PHACTR2* in first trimester informative trophoblasts.

So, rs1082 and rs2073214 show partial imprinting in human trophoblast. The imprinting seems more pronounced in term placenta than at the beginning of the pregnancy. This could be due to differential *PHACTR2*-isoform regulation throughout gestation.

5.1.3. Discussion.

We found *PHACTR2* to be partially imprinted in term and first trimester human placenta. *PHACTR2* is located on chromosome 6q24.2, 114 kb from *PLAGL1* a known imprinted gene (previously called *ZAC*) (KAMIYA *et al.* 2000). *PLAGL1*

is paternally expressed and maternally methylated (ARIMA *et al.* 2000). As loss of imprinting of *PLAGL1* is seen in TNDM (GARDNER *et al.* 2000; MA *et al.* 2004), genes in the locus have previously been inspected for imprinted expression and *PHACTR2* (*KIAA0680*) was previously tested on monochromosomal hybrid cells (human chromosome donor fibroblast, mouse recipient cells and hybrids containing a paternal or maternal copy of chromosome 6) but biallelic expression was reported (ARIMA *et al.* 2000). We extensively tested *PHACTR2* by Sanger sequencing and demonstrated partial imprinting with expression biased towards the maternal allele.

PHACTR2 is a member of a family of four actin and protein phosphatase 1 (PP1) binding proteins highly expressed in the brain (SAGARA *et al.* 2003; ALLEN *et al.* 2004). Little is known about the function of the *PHACTR* genes family (*PHACTR1-4*). *PHACTR2* (phosphatase and actin regulator 2) is expressed in various tissues and highly expressed in nerve and placenta (Unigene database at http://www.ncbi.nlm.nih.gov/UniGene/). The function of *PHACTR2* in placenta is unknown at present.

A two tier GWAS has found rs11155313 in *PHACTR2* to be significantly associated with Parkinson's disease (PD) in a replication study on a series of matched unrelated patient-control pairs (MARAGANORE *et al.* 2005). The association of the same SNP with PD was studied in four Caucasian patient-control series (from the US, Ireland, Norway and Canada). The association was significant in the Irish, American and Norwegian series (WIDER *et al.* 2009). *PHACTR2* has also been recently implicated in multiple sclerosis (GOLDSTEIN *et al.* 2010).

Recently, RNA-seq of mouse placenta cDNA samples of F1 reciprocal crosses between AKR and PWD strains detected partial imprinted expression of *Phactr2* (WANG *et al.* 2011). They used Illumina Genome Analyzer to sequence total RNA samples from E17.5 placentas of the reciprocal crosses. They found 251 candidate genes (q-value <0.01) including 35 genes that were already known to be imprinted. Most known imprinted genes in their screen generated the highest

q-value rank. To detect significant parent-of-origin expression ratio, they defined p_1 as the percentage of maternal allele in AKR female x PWD male cross and p_2 as the paternal percentage for PWD x AKR. *Phactr2* was highlighted as being partially imprinted by the RNA-Seq experiment with p_1 66% and p_2 34%. These percentages are similar to what we observe on the sequence tracks obtained (Figures 44, 45 and 46). Allele-specific expression was confirmed in five out of seven candidates by pyrosequencing in multiple placental samples of the same crosses. For *Phactr2*, the biased expression towards the maternal allele was confirmed (WANG *et al.* 2011).

In summary, our results show that *PHACTR2* is partially imprinted in human placenta and recent work by another group show it is also partially imprinted in mouse placenta (WANG *et al.* 2011). The imprinted expression is thus conserved between mouse and human.

It would be interesting to test whether *PHACTR2* is co-regulated with *PLAGL1/HYMA1*. It would be useful to find patients with mutations, small deletions or localised duplications that disrupt it. Mouse models or human CNV models would also be helpful.

5.2. Confirmation *ZNF331* imprinting and identification of two DMRs.

5.2.1. Introduction.

The zinc finger 331 gene (*ZNF331*; CCDS 33102.1) maps to hChr19q13.42. Using two SNPs, rs8100247 (exon 1, 5'UTR) and rs12982082 (exon 2, 5'UTR) on the Illumina system, *ZNF331* exhibited a consistent pattern of maternal origin for the expressed allele (Figure 35 in previous Chapter).

This gene was first identified as a candidate disrupted gene in translocations involving 19q13 in thyroid adenomas and named Rearranged in Thyroid Adenomas (*RITA*) (RIPPE *et al.* 1999). It was also named ZNF463 as it encodes a 463 amino acids protein with an amino-terminal KRAB (Krüppel associated box) domain and 12 carboxy-terminal C_2H_2 zinc finger units (UniProt Q9NQX6) (WU *et al.* 2001). Its name was later changed to *ZNF331* according to the HUGO Gene Nomenclature (http://www.genenames.org/). Three transcripts in RefSeq (NM_001079906, NM_001079907, and NM_018555) correspond to proteins NP_001073375.1, NP_001073376.1 and NP_061025.5. Meiboom et al. performed rapid amplification of cDNA ends to characterize the longer transcripts and aligned the cDNA sequences with gDNA to predict the TSS (MEIBOOM *et al.* 2003). More recently, the ENCODE project identified functional elements by many different technologies and these data can be browsed in UCSC genome browser (Figure 47).

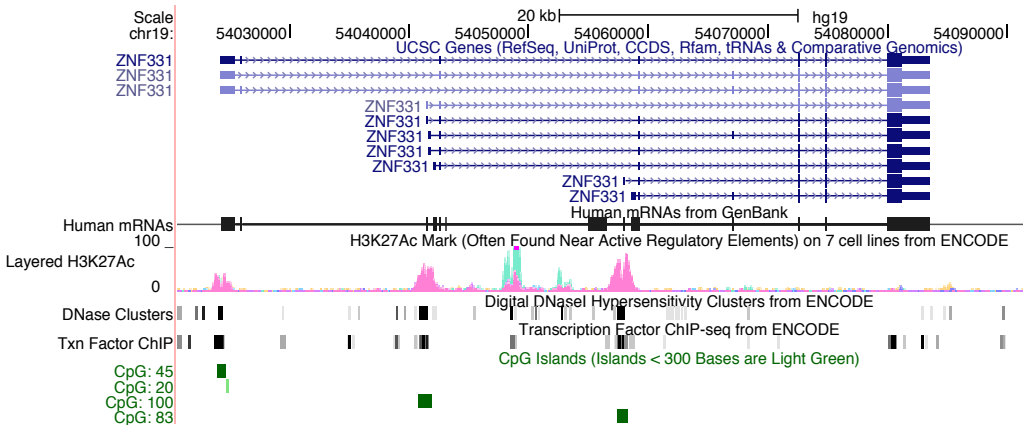

Figure 47: *ZNF331* pictogram in UCSC genome browser showing ENCODE regulatory tracks.

Based on the transcription factor binding sites ChIP data (TXn Factor), there are three promoters and a likely fourth one, in between the second and third (Figure 47).

The mRNA levels are high in a range of normal tissues including embryonic tissue according to the UniGene expression database (http://www.ncbi.nlm.nih.gov/unigene). In the Novartis GNF gene expression atlas (BioGPS), the expression is higher in the pituitary, ovary and adrenal gland compared to other tissues (http://biogps.gnf.org/#goto=genereport&id=55422). In the ArrayExpress database, the mRNA levels were increased in term placenta compared to midgestation in human (experiment deposited in Gene Expression Omnibus data repository with accession no. GSE5999) (Figure 48) (WINN *et al.* 2007).

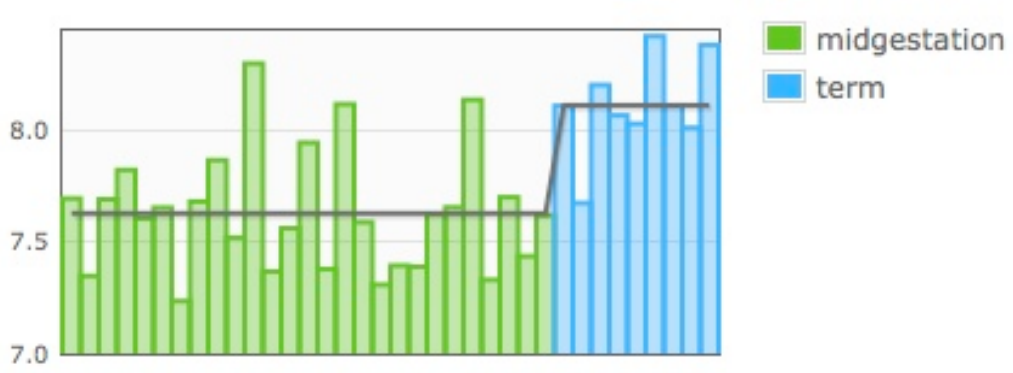

Figure 48: ArrayExpress pictogram of *ZNF331* expression in placenta throughout gestation.
Microarray data generated and deposited by Winn et al. (WINN *et al.* 2007)

Human *ZNF331* gene was first discovered to be imprinted in LCLs by Kelly Frazer's group in screens using oligonucleotide arrays (Perlegen) (PANT *et al.* 2006; POLLARD *et al.* 2008). In their first study, *ZNF331* (named *ZNF463*) was *maternally* expressed in LCLs from five children in a single CEPH pedigree using rs8109631 (ss24225694 on their Perlegen array) and two additional SNPs in the 3' exon. In the 5' exon, they used the same SNP as in this study, rs8100247 (ss24225691), and showed biallelic expression in LCLs (PANT *et al.* 2006). In their second study, *ZNF331* was studied in another informative CEPH pedigree using again rs8100247. Quantitative sequencing was used to confirm the microarray data. In this second study, *ZNF331* transcripts targeted by rs8100247 were reported to be *paternally* expressed (POLLARD *et al.* 2008) (see Table 17 for summary). The opposing parental orientation of imprinting was not discussed (POLLARD *et al.* 2008). The discrepancies in the allelic expression pattern are difficult to explain, but apart from a technical error, could be due to isoform specific imprinting.

Table 17: Summary of *ZNF331* SNPs tested in LCLs by Pant et al. and Pollard and al. We note that the rs8100247 and rs8109631 results are discordant.

Pant et al.	Oligonucleotide microarray	12 unrelated individuals	Real-time PCR, CEPH 1362 pedigree
SNP	**Location on chr19**	**Allelic expression pattern**	**Additional data**
rs8100247	5' UTR, exon 1, 58717013	Biallelic	
rs8100455	5' UTR, exon 1, 58717100	Biallelic	
rs8100338	5' UTR, exon 1, 58717006	Biallelic	
rs16985052	3', exon 7 CDS, 58772592	ASE in LCLs	
rs1056393	3', exon 7 CDS, 58772811	ASE in LCLs	
rs8109631	3', exon 7 CDS, 58771956	ASE in LCLs	5 htz indiv., maternal expression LCLs

Pollard et al.	Oligonucleotide microarray	67 unrelated individuals, 3 ethnic groups	Quantitative sequencing, CEPH 1444 pedigree
SNP	**Location on chr19**	**Allelic expression pattern**	**Additional data**
rs8100247	5' UTR, exon 1, 58717013		Imprinted, paternal expression LCLs
rs16984961	5' UTR, exon 1, 58717090		Sequencing in osteoblast-like cell lines ASE 2.6 fold
rs1056393	3', exon 7 CDS, 58772811	ASE	
rs16985052	3', exon 7 CDS, 58772592	ASE	
rs1351	3', exon 7 CDS, 58772091		Sequencing in osteoblast-like cell lines ASE 22-fold

5.2.2. Confirmation of imprinting by Sanger sequencing.

To confirm the imprinting and maternal expression of *ZNF331* discovered in placenta in this study, RT-PCR amplification and Sanger sequencing of two SNPs in two exons of the *ZNF331* transcript (exon 1, 5'UTR and exon 7, CDS) were carried out. The rs8109631 SNP is located in exon 7, which is shared by all transcripts while rs8100247 was used to test the longer transcripts (Figure 49).

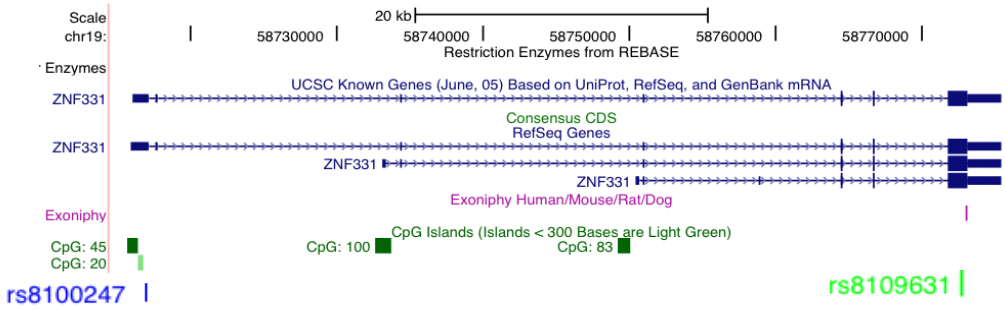

Figure 49: Location of markers SNPs in *ZNF331*.
Figure generated with UCSC genome browser.

5.2.2.1. ZNF331 imprinting status in term placenta.

Primers were designed around rs8100247, with the sense primer in exon 1 and the antisense primer in exon 2 of *ZNF331* (Figure 49 and 50). Seven informative placental samples were sequenced using different sets of primers (listed in Appendices section) for gDNA and cDNA. The sequence traces overlapping rs8100247 confirmed the exclusive maternal expression of this exon seen with the Illumina method.

Three informative samples were sequenced with rs8109631 as marker. While it was always the maternal allele that was more expressed, there was significant expression of the paternal allele. These data suggested that at least one 3' isoform of *ZNF331* is not perfectly imprinted.

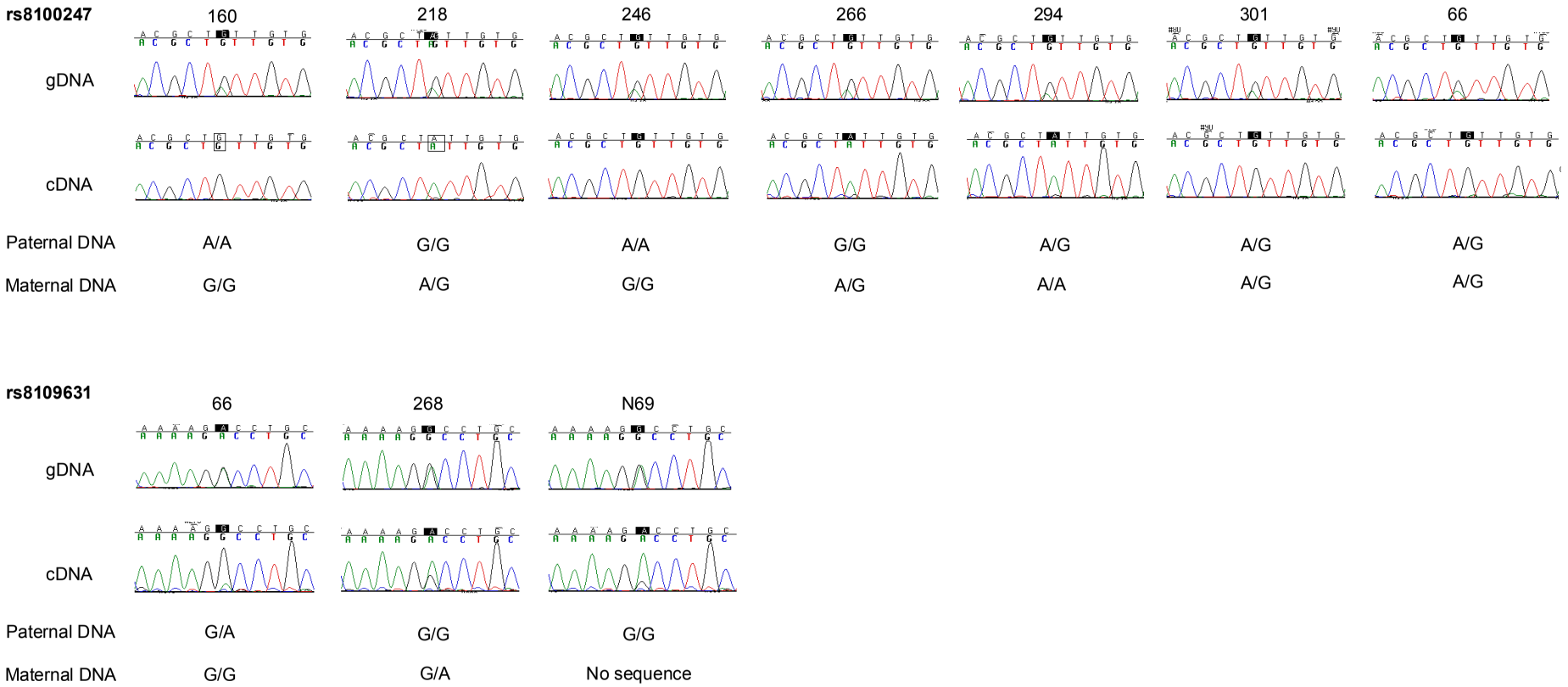

Figure 50: Confirmation of *ZNF331* imprinting in term placenta using Sanger sequencing.

Sequences for rs8100247 (exon 1, 5' UTR) (top) and rs8109631 (last exon, CDS) (bottom) of informative term placenta samples in gDNA (top trace for each set) and cDNA (bottom trace) with corresponding Illumina genotyping data for the parents. There were seven informative samples available for analysis for rs8100247 and three for rs8109631.

5.2.2.2. ZNF331 imprinting status in first trimester placenta.

ZNF331 imprinting status was also examined in first trimester placentas by Sanger sequencing. The sequences obtained for the same marker SNPs are presented in Figure 51 (see next page) and confirm the monoallelic expression of at least some isoforms. The imprinting is in the same parent-of-origin direction (maternal expression) as in the term placenta.

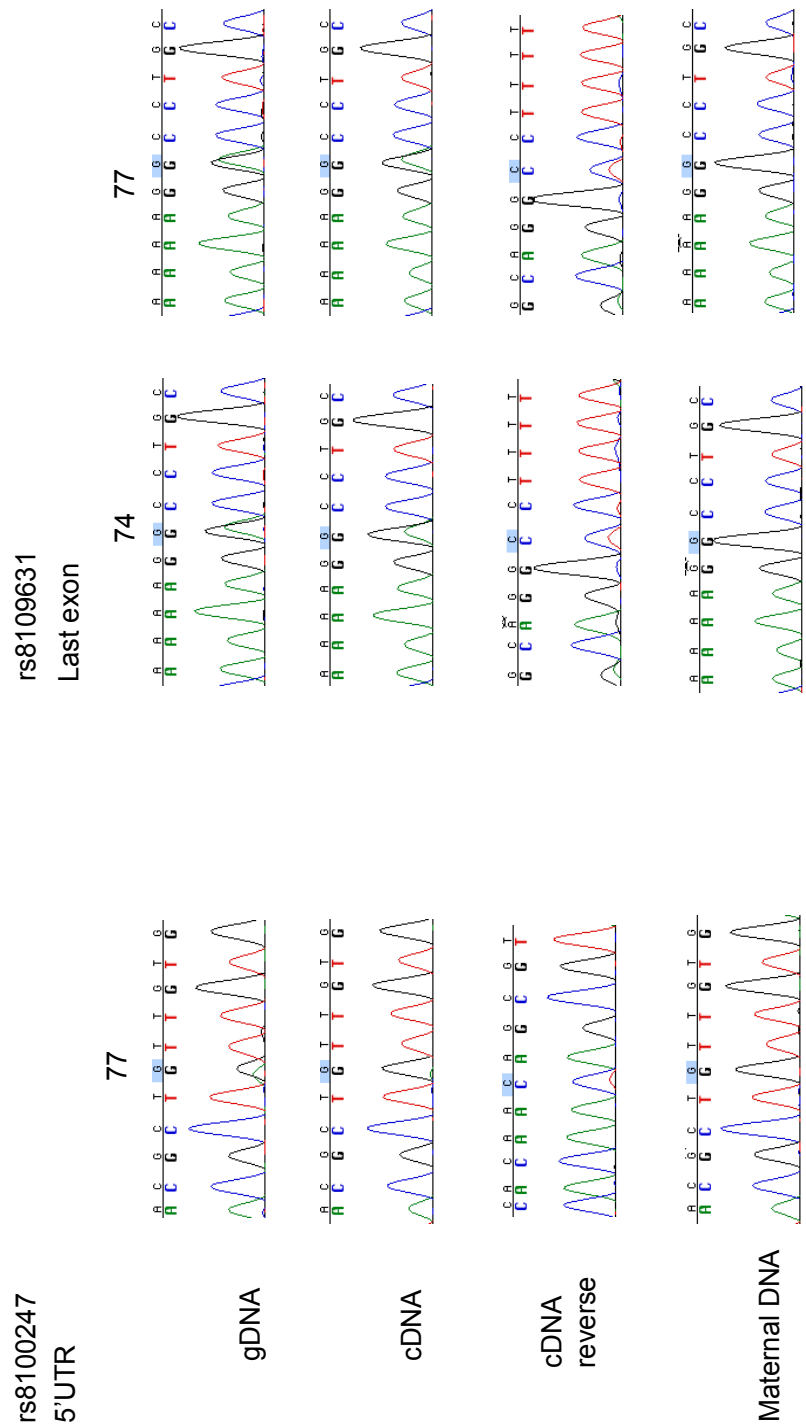

Figure 51: *ZNF331* imprinting status in first trimester placenta.

Sequences for the same marker SNPs confirm imprinting of at least some isoforms in the same parent-of-origin direction at this developmental stage; 14 weeks of gestation for sample 77 and 11 weeks for sample 74.

5.2.2.3. ZNF331 isoform specific imprinting.

Exploratory work to determine the imprinting status of the putative *ZNF331* isoforms was started by Sanger sequencing. According to the UCSC genome browser, there are three CpG islands overlapping the putative promoters of *ZNF331* (CpG 45, 100 and 83) (Figure 52) and a small additional one adjacent to the first 5' CpG island (CpG 20).

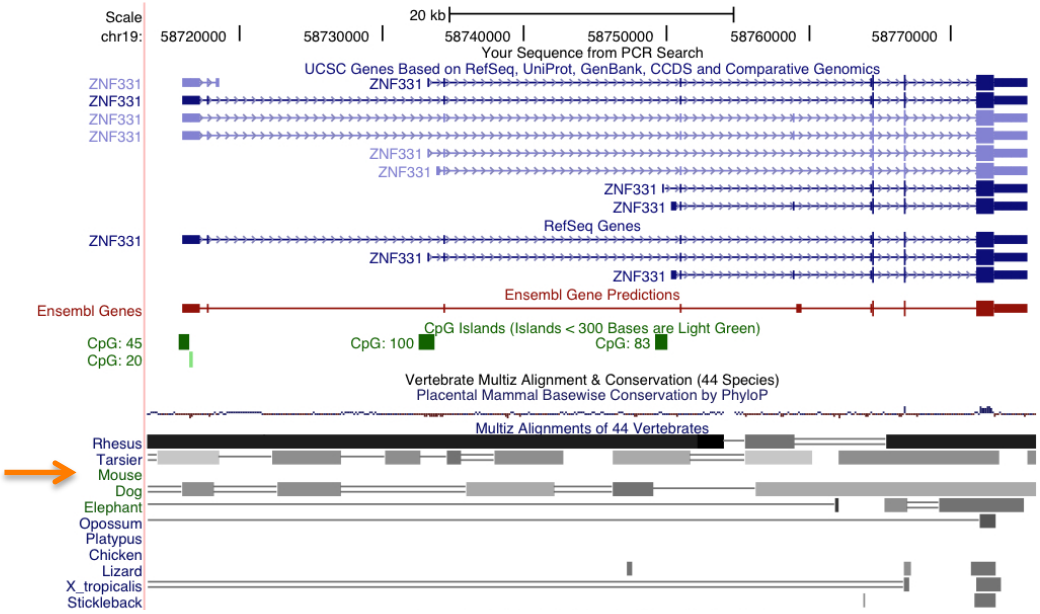

Figure 52: Putative transcripts of *ZNF331* and conservation across species.
Figure generated with UCSC genome browser. The bottom tracks on the Figure represent the conservation of the gene. There is no described mouse orthologue for *ZNF331* (green arrow).

We first attempted to decipher *ZNF331* imprinting by exon specific RT-PCR. Our results show that the longest isoforms, starting at 'CpG45', are maternally expressed (Figure 53). Due to technical issues and lack of informative SNPs, we were not able to elucidate which 3' isoform is not or only partially imprinted. So we turned our attention to the DMRs as these are often indicators of imprinting status in other imprinted genes.

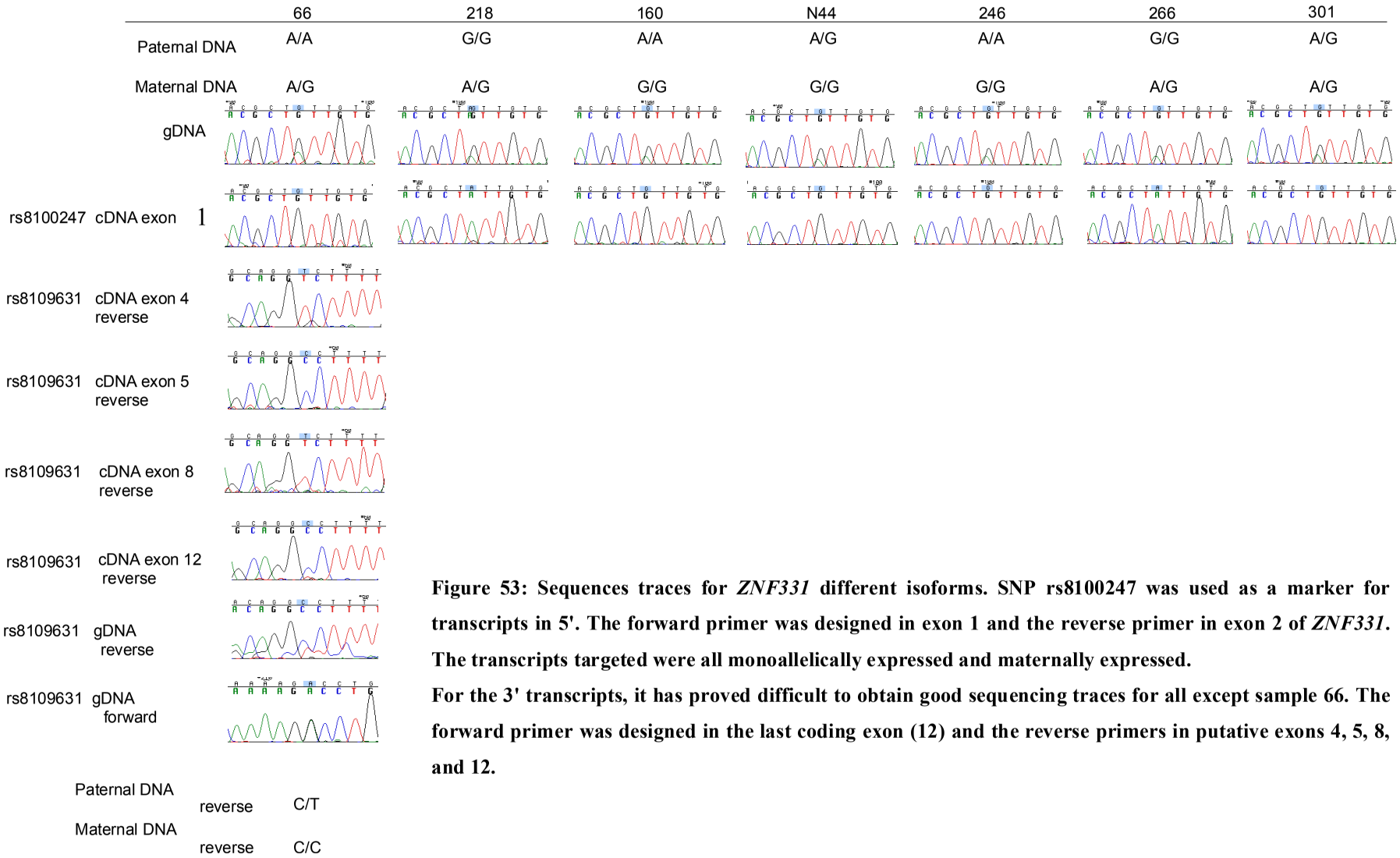

Figure 53: Sequences traces for *ZNF331* different isoforms. SNP rs8100247 was used as a marker for transcripts in 5'. The forward primer was designed in exon 1 and the reverse primer in exon 2 of *ZNF331*. The transcripts targeted were all monoallelically expressed and maternally expressed.

For the 3' transcripts, it has proved difficult to obtain good sequencing traces for all except sample 66. The forward primer was designed in the last coding exon (12) and the reverse primers in putative exons 4, 5, 8, and 12.

5.2.3. *ZNF331* DMR identification.

In each imprinted locus there is a requirement for a germline DMR. According to UCSC genome browser, there are four CpG islands overlapping *ZNF331* putative promoters (Figure 52 and Table 18).

Table 18: Location of *ZNF331*-CpG islands according to UCSC genome browser. The table lists their start and end sites and their size.

UCSC (hg 19)	Start chr 19	End chr19	Size
CpG 45	54023869	54024560	692
CpG 20	54024646	54024923	278
CpG 100	54040813	54041857	1045
CpG 83	54057415	54058254	840

5.2.3.1. Combined bisulfite and restriction analysis (CoBRA) of CpG100.

CoBRA and bisulfite sequencing were used to screen these four CpG islands for differential methylation. Only the CoBRA analysis of CpG100 (54040813-54041857) suggested differential methylation of the alleles (Figure 54). Digested and undigested products were detected which suggested the presence of both methylated and unmethylated DNA (third lane in Figure 54). CpG100 is the CpG island of the putative second promoter and is located upstream of exon 3.

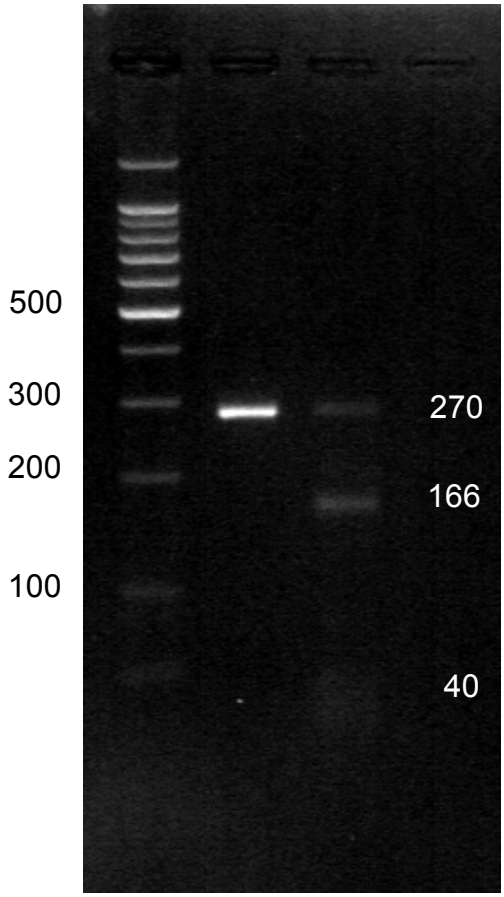

Figure 54: Gel electrophoresis of restriction of CpG100.
The PCR product of bisulfite treated placental DNA appeared partially digested which was suggestive of partial methylation. The DNA ladder in the first lane enables the estimation of fragments size. The unrestricted PCR product was loaded on the gel in the second lane for comparison. In the third lane, fragments appear suggesting differential methylation of the alleles.

5.2.3.2. Sequencing of placental DNA treated with bisulfite.

The methylation pattern of these CpG islands plus the next one telomerically was also more broadly studied by sequencing of gDNA treated by bisulfite. The data was analysed with the BDPC webtool (http://biochem.jacobs-university.de/BDPC/) (ROHDE *et al.* 2008) (see section 2.10 of Chapter 2). For CpG100 on 34 clones sequenced, 66% of CpGs were methylated and 33% were unmethylated (Figure 55). All clones were unmethylated for CpG45 and all were methylated for CpG83. CpG20 was not studied separately from CpG45. Bisulfite

sequencing confirmed that differential methylation was present for CpG island 100. Unfortunately, no SNP was present in CpG100 bisulphite clones and we could not determine whether CpG100 is a parent-specific DMR or an allele-specific DMR.

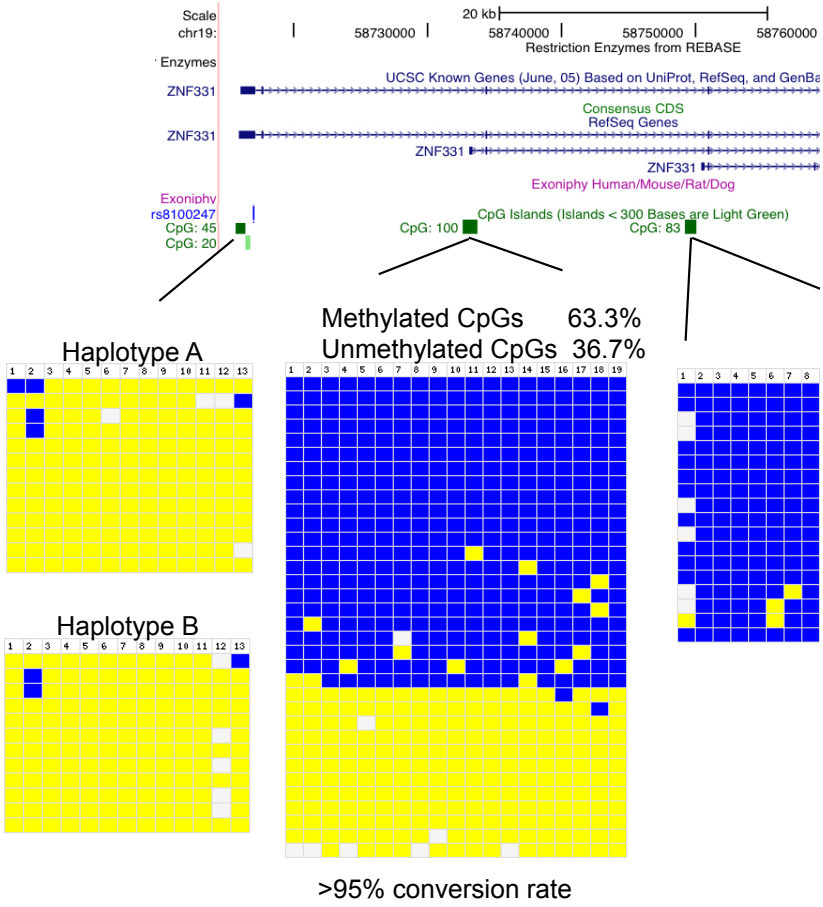

Figure 55: DNA methylation levels of CpG islands within *ZNF331*.

The figure shows the chromosomal location of *ZNF331*, the SNPs used for sequencing and the CpG islands in the UCSC genome browser. In the bottom part of the figure, the methylation patterns of the three CpG islands studied are represented after processing the data using the BDPC webtool. The blue colour indicates methylated CpGs and the yellow colour unmethylated CpGs. Each column represents a single CpG site and each row represents one clone. The sequences for CpG 45 are sorted according to their haplotype.

The next CpG island telomerically (CpG86) that lies at chr19:58842889-58844053 between the promoter of the *DPRX* gene and the C19MC-microRNA cluster was also studied by bisulfite sequencing (Figure 56). This study demonstrated differential methylation of this CpG island as well (74% of methylated CpGs and 26% of unmethylated ones) (Figure 56).

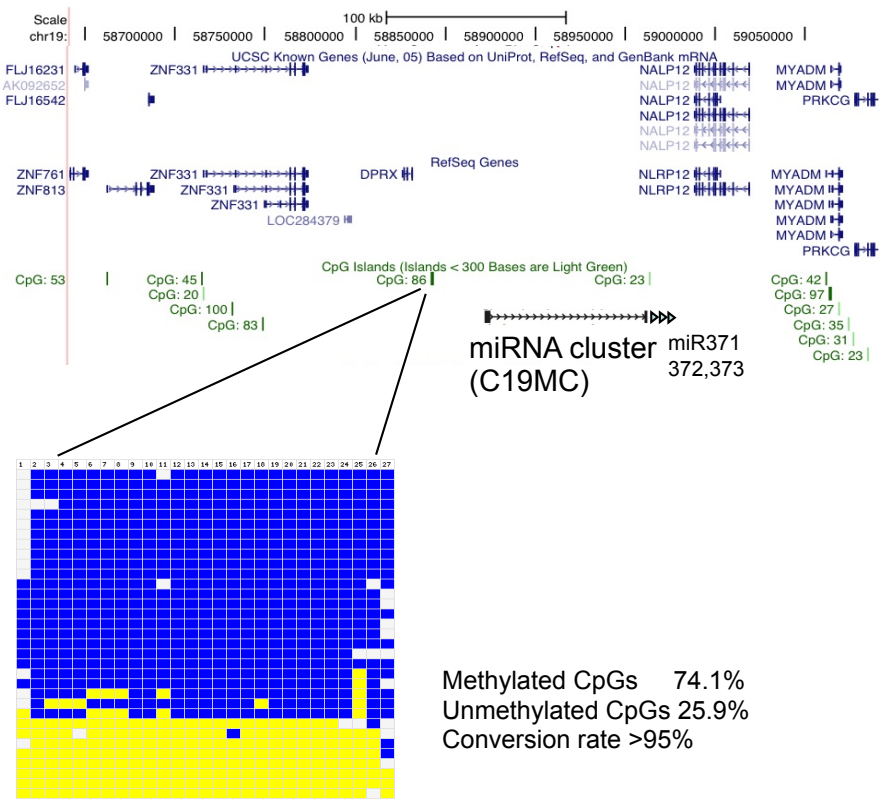

Figure 56: Methylation level of CpG86 in *ZNF331* locus.

The figure, generated with UCSC genome browser, indicates the position of CpG island in relation to *ZNF331* location, neighbouring genes and miRNAs. Bisulphite sequencing data was compiled with the BDPC webtool as in Figure 55. No SNP was included in the sequence.

5.2.4. Discussion of *ZNF331* study.

These data confirm that *ZNF331* is imprinted in human term placenta and that it is expressed from the maternal allele. The longer transcripts of *ZNF331* exhibit complete imprinting when tested with rs8100247 (located in exon 1 and common to several transcripts in 5') using the ASE Array and Sanger sequencing. When testing all transcripts pooled at the 3' UTR (rs8109631), the allelic expression was biased towards the maternal allele (Figure 57). Even though the number of samples was small, the sequence traces of RT-PCR products of first trimester placentas corroborated the orientation of imprinting and the incomplete imprinting of all transcripts pooled at the 3' UTR.

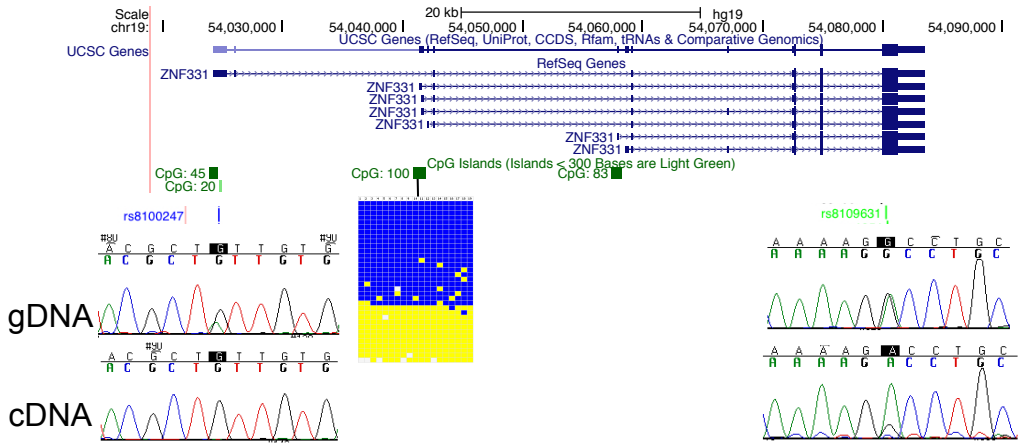

Figure 57: *ZNF331* study summary.

The sequence traces exemplify the data obtained for the two marker SNPs (see Figure 50 for exhaustive results). The sequencing data of bisulfite treated gDNA is shown for CpG100.

ZNF331 (also known as *ZNF463*) was first shown to exhibit monoallelic expression in a parent-of-origin manner in LCLs (PANT *et al.* 2006; POLLARD *et al.* 2008), although the parent-of-origin orientation of *ZNF331* in these studies was not clear. These findings prompted us to study each previously known transcript separately but this has proved difficult due to their varying mRNA expression levels in the placenta.

We have identified two DMRs in the locus: CpG100 (methylated clones 66% and unmethylated clones 33%) and CpG86 (methylated clones 74% and unmethylated clones 26%). CpG100 overlaps with a putative promoter of *ZNF331* and was differentially methylated in the same proportions in the study by Harris et al. (Figure 58) (HARRIS *et al.* 2010).

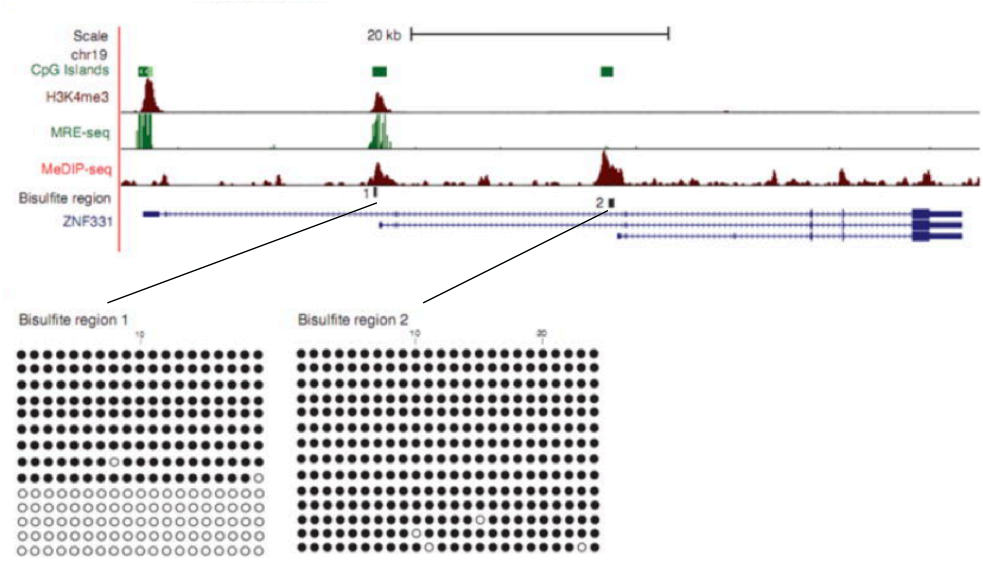

Figure 58: Confirmation of *ZNF331* DMR by another group.
Clonal bisulphite sequencing results obtained by Harris et al. (HARRIS *et al.* 2010) confirm the differential methylation at CpG100 (labelled '1' in their figure) and full methylation of CpG83 (labelled '2'). Figure reused with permission.

Recently, another new high-throughput technology named ChIP-SNP has been used to map allele-specific binding of RNA polymerase II (RNAP) in the human fibroblast genome (MAYNARD *et al.* 2008). This method combines a modified ChIP-chip and SNP genotyping arrays (HumanHap300, Illumina). A subset of 466 SNPs (4%) showed allele-specific enrichment by RNAP ChIP. These SNPs corresponded to 239 RefSeq genes, 2 small nucleolar RNAs and 16 microRNAs. There were five known imprinted genes in these genes (*SNRPN*, *KCNQ1*, *PLAGL1*, *HYMAI*, *MEG3* (also known as *GTL2*) and its adjacent microRNA cluster). Two marker SNPs, that were of interest to us, were shown to be subjected to allele-specific binding: rs4803143, which is located in the second intron of *ZNF331*, and rs1293700, which is 5374bp distant from CpG island 86

that we found differentially methylated in the *ZNF331* locus. These data suggest allele-specific transcriptional repression in the *ZNF331* locus (MAYNARD *et al.* 2008).

Another recent study tested allele-specific methylation across the genome. Two SNPs in *ZNF331* showed evidence for allele-specific methylation: rs8105870, which is 281 bp away from CpG100 and rs7248353, which is 259 bp away from CpG83 (SCHALKWYK *et al.* 2010).

The next CpG island that we studied (CpG86) is located between the promoters of *DPRX* and *C19MC* and has also been shown to be differentially methylated by another group in the same proportions (Figure 59) (TSAI *et al.* 2009).

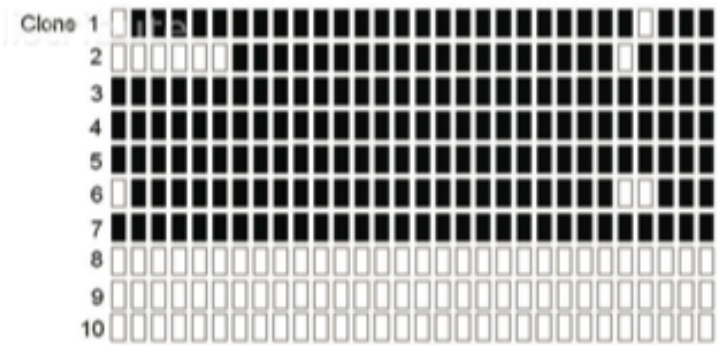

Figure 59: Bisulphite sequencing of CpG86.
Data obtained by Tsai et al. in Placenta (TSAI *et al.* 2009). Figure reused with permission.

ZNF331 is thus imprinted in human placenta in addition to lymphocytes and fibroblasts suggesting ubiquitous imprinting. Such ubiquity calls for the study of its function in relation to metabolism, behaviour, fetal development and cancer.

5.3. Exploration of the '*ZNF331* locus'

Imprinted genes are often found in clusters throughout the genome (THORVALDSEN and BARTOLOMEI 2007) and share regulatory elements (VERONA *et al.* 2003). No imprinted gene was known in the vicinity of *ZNF331*. The exploration of the locus was started in 3' of ZNF331, where *LOC284379* (solute carrier 7) gene encoding a cationic amino acid transporter that is specifically expressed in the bladder (in the Unigene database) lies (Figure 60). Further downstream lies *DPRX*, a divergent-paired homeobox gene. Homeobox gene families have a conserved DNA motif in common: the homeodomain. Homeodomain proteins are thought to function as transcription factors that are important for embryonic development (BOOTH and HOLLAND 2007). Seven retrotransposed copies of *DPRX* exist. It has no identified mouse orthologue and its rate of sequence evolution is accelerated suggesting a possible role in human reproduction biology (BOOTH and HOLLAND 2007). There is no expression data for *DPRX* in the Unigene database or BioGPS and no suitable SNP was found to study its allelic expression using the UCSC genome browser (Figure 60, see next page).

Further telomerically, *NLRP12* encodes for NLRP12 protein (NP_653288.1), which is a CATERPILLER protein (Figure 60). The *NLRP* (Nucleotide-binding oligomerization domain, Leucin rich repeat and Pyrin domain) gene family consists in 14 members encoding proteins with a similar structure and divided into two gene clusters. One cluster is located at 11p15 (*NLRP6*, *10* and *14*) and the other at 19 q13.4 (*NLRP2, 4, 5, 7, 8, 9, 11, 12*, and *13*).

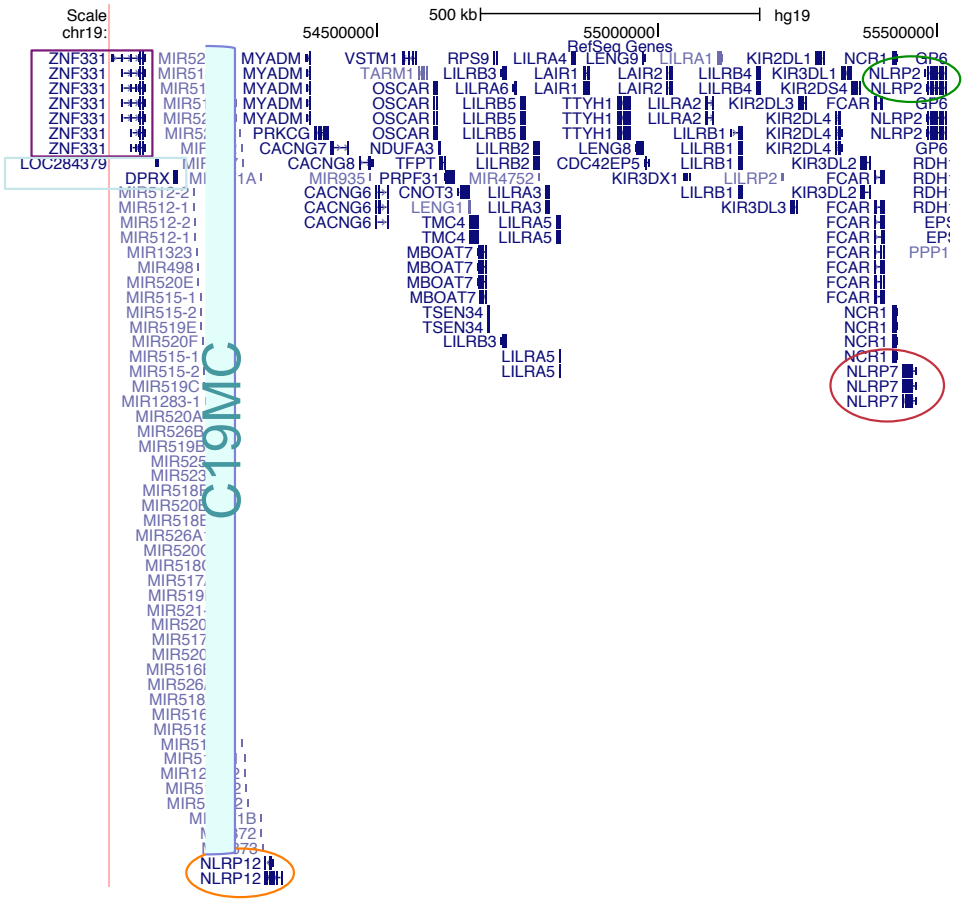

Figure 60: *ZNF331-NLRP2* interval on chromosome 19q13.42.

NLRP12 (*NALP12*) is implicated in familial cold autoinflammatory syndrome 2 (FCAS2), a condition characterised by episodes of fever, urticaria, myalgias, and arthralgias following exposure to cold. The pedigrees of two unrelated families with FCAS2 in Guadeloupe were consistent with autosomal dominant inheritance (JERU *et al.* 2008). Several other *NLRP* genes have been associated with human disease and imprinting (Table 19). Interestingly, provisional data have suggested that *NLRP2* (55.477.711-55.512.510 on hChr19 further telomerically) is partially imprinted with expression skewed towards the maternal allele in human placentas (BJORNSSON *et al.* 2008). A mutation in *NLRP2* gene has also been implicated in familial Beckwith-Wiedemann syndrome and *NLRP2* seems to have a role in imprinting by *trans* regulation (MEYER *et al.* 2009). For its neighbour *NLRP7*, a role in *trans* regulation of imprinting has been established. Female homozygotes fail to methylate DMRs at several imprinted loci in their germ cells, which leads to biparental hydatiform

moles on a recurrent basis (DJURIC *et al.* 2006; MURDOCH *et al.* 2006; KOU *et al.* 2008).

Table 19: Syndromes associated with NLRP genes.
The gene is listed with the disease caused by its dysregulation, the entrance number in the OMIM database and the chromosomal location. *NLRP 12*, *7* and *2* are located at 19q13.42.

Gene	Disease	MIM	Chr location
NLRP12	Familial cold autoinflammatory syndrome 2 (JERU *et al.* 2008)	611762	19q13.42
NLRP7	Familial recurrent hydatiform moles, recurrent miscarriages, intra-uterine growth restriction, stillbirths (MURDOCH *et al.* 2006)	231090	19q13.42
NLRP2	Familial imprinting disorder (Beckwith-Wiedemann Syndrome) (MEYER *et al.* 2009) Preliminary imprinting data (maternal expression) (BJORNSSON *et al.* 2008)	Not listed	19q13.42
NLRP3	Chronic infantile neurological cutaneous and articular syndrome (CINCA syndrome) (FELDMANN *et al.* 2002)	606416	1q44
NLRP1	Vitiligo-associated multiple autoimmune disease susceptibility 1 (VAMAS1) (JIN *et al.* 2007)	606579	17p13.2

NLRP12 is however more than 1 Mb distant from *NLRP2* and *NLRP7* that lie next to each other. Unfortunately, the *NLRP12* gene in the *ZNF331* locus could not be tested for imprinting because it was not sufficiently expressed in the placenta (S. Abu-Amero unpublished data).

Between *DPRX* and *NLRP12*, lies the longest microRNA (miRNA) cluster found in human that extends over approximately 100 kb (BENTWICH *et al.* 2005). This cluster is located on chr19 at position 58,861,745-58,961,404 and has been named *C19MC* (Chr19 miRNA cluster) (Figure 61) (BORCHERT *et al.* 2006). It

comprises 54 predicted microRNA genes of which 43 have been sequenced (BENTWICH *et al.* 2005). It is does not seem to be conserved beyond primates and is almost exclusively expressed in placenta (BENTWICH *et al.* 2005; BEREZIKOV *et al.* 2006). Adjacent to this cluster lies another cluster (mir-371, 2 and 3), which is conserved in dog and mouse (BENTWICH *et al.* 2005; BORTOLIN-CAVAILLE *et al.* 2009) (Figure 66).

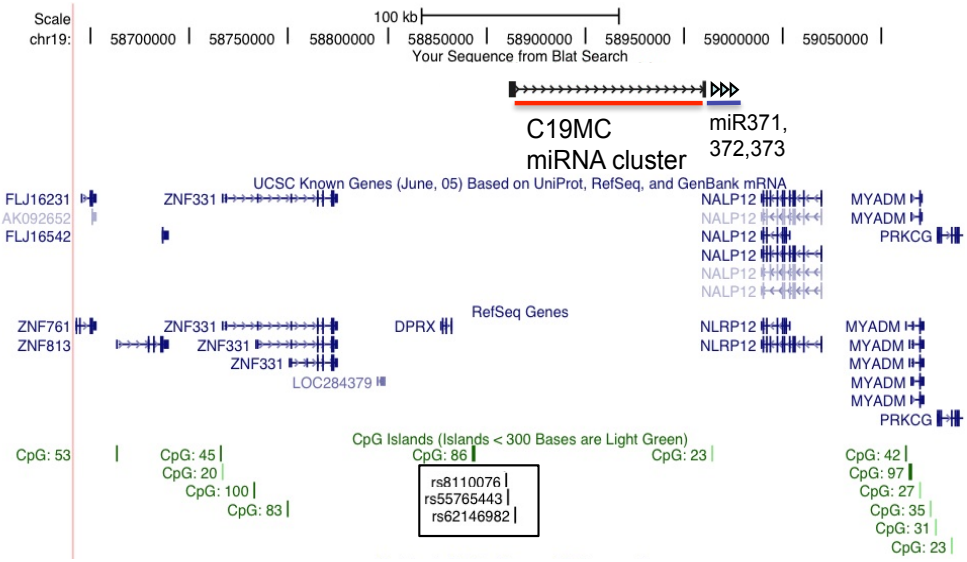

Figure 61: *ZNF331* locus and CpG islands.
Location of *ZNF331*, *LOC284379*, *DPRX*, *C19MC* miRNA cluster, miR371, 372 and 373 are depicted in UCSC. The three SNPs used as markers to test the imprinting status of *C19MC* are shown (rs8110076, rs55765443 and rs62146982).

5.3.1.1. Imprinting of the C19MC pre-miRNA.

Several SNPs were available so it was possible to test the imprinting status of the pre-miRNA *C19MC* (Figure 61). To do this, the cDNA of informative first trimester trophoblast samples was sequenced. The results showed partial imprinting of *C19MC* with expression from the paternal allele (Figure 62).

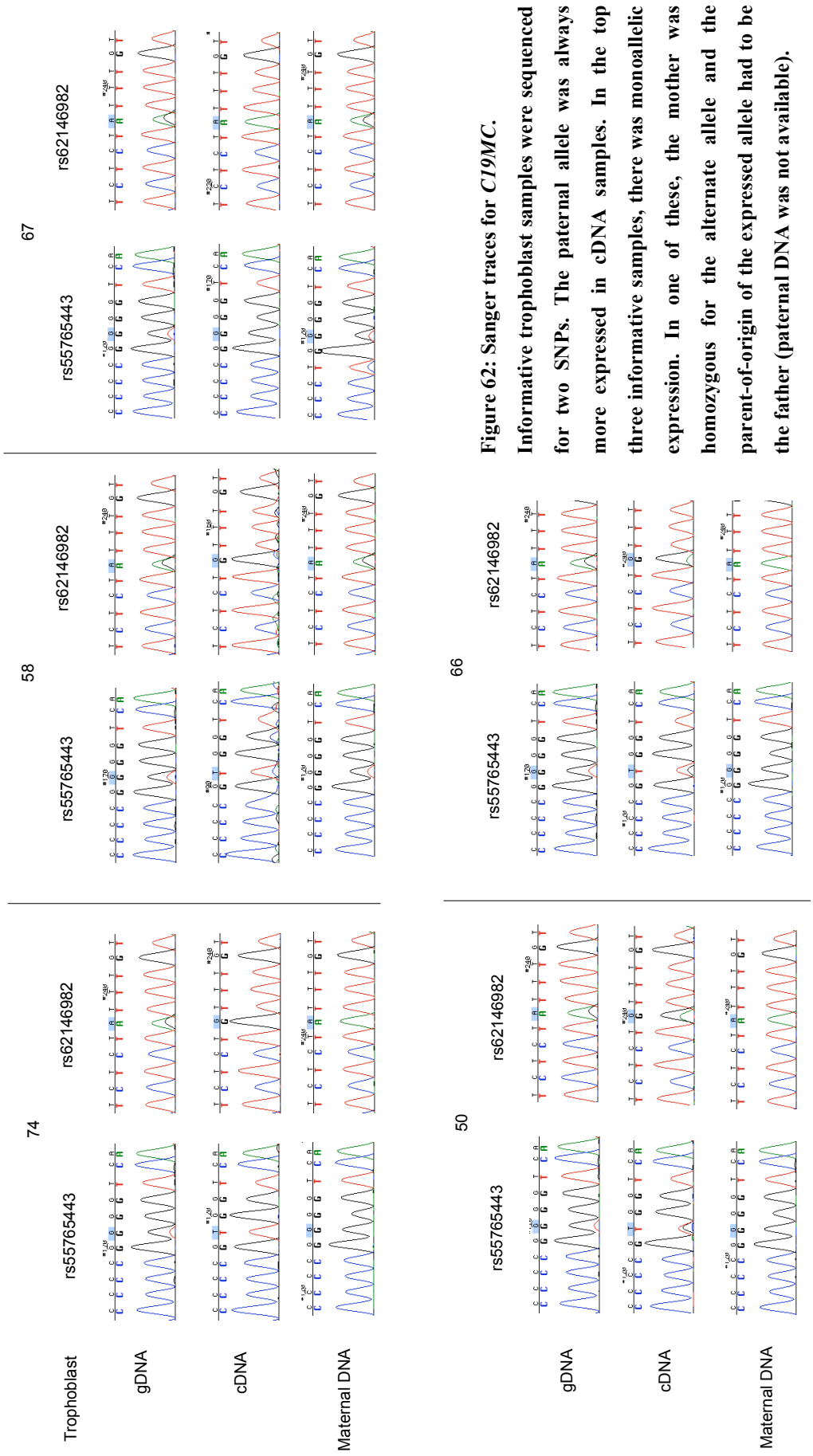

Figure 62: Sanger traces for *C19MC*.
Informative trophoblast samples were sequenced for two SNPs. The paternal allele was always more expressed in cDNA samples. In the top three informative samples, there was monoallelic expression. In one of these, the mother was homozygous for the alternate allele and the parent-of-origin of the expressed allele had to be the father (paternal DNA was not available).

5.3.2. Discussion of the exploratory work on the *ZNF331* locus.

The large primate-specific microRNA cluster *C19MC* was studied in first trimester placentas and was partially imprinted at this stage of gestation. Prof. Moore's laboratory in collaboration with Jérôme Cavaillé has since demonstrated *C19MC* to be fully imprinted in 22 informative term placentas (NOGUER-DANCE *et al.* 2010). The expressed allele was confirmed to be the paternal one. Noguer-Dance et al. also demonstrated that CpG86 is a germline DMR (NOGUER-DANCE *et al.* 2010).

The current organisation of this new human imprinted locus can be summarised as in Figure 63. The locus contains at least two placental partially imprinted genes (*ZNF331* and *C19MC*) and at least two DMRs (CpG100 and CpG86).

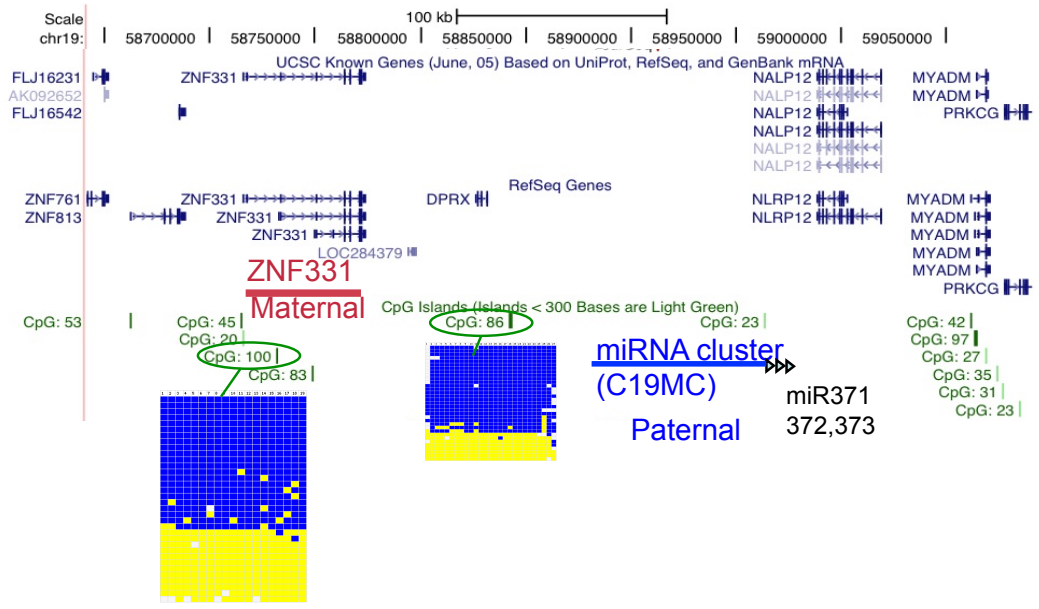

Figure 63: Summary of *ZNF331* locus organisation.
***ZNF331* (in red) is imprinted and maternally expressed. *C19MC* (in blue) is imprinted and paternally expressed. Circled CpG islands are the two DMRs. The imprinting status of the other genes in the locus is unknown.**

To further elucidate this imprinted locus, detailed characterisation of its transcriptome by bioinformatics and functional tests would be necessary.
Being a primate-specific imprinted locus, future studies on *ZNF331-C19MC* should be interesting for an evolutionary as well as a molecular, physiological and medical point-of-view.

Chapter 6. Discussion.

6.1. Summary.

This thesis has described a screen of human term placenta using two different technologies that enable the quantitation of alleles at single nucleotide variations. Trios consisting of human placenta tissue samples along with gDNA from mother and father permitted the analysis of imprinted versus non-imprinted ASE in selected candidate genes. The Sequenom technology was used to test 131 candidates across a core set of 23 trios (Chapter 3). The second technology from Illumina increased the throughput significantly (1536 SNPs in 932 genes) and ASE was detected in 18% of genes tested in at least two placental samples (Chapter 4). Two new examples of partial imprinting were found. One was a novel, partially imprinted gene (*PHACTR2*), which extends an imprinted cluster that is physically linked to a known imprinted gene (*PLAGL1*). The other gene was already known to be imprinted in another tissue but had not been tested in placenta (*ZNF331*) (Chapter 4). *ZNF331* partial imprinting was confirmed by Sanger sequencing on additional samples and also in placenta during the first trimester of pregnancy. *ZNF331* is maternally expressed while the microRNA gene cluster (*C19MC*), which lies adjacent is also imprinted but paternally expressed. This new imprinted locus on hChr19 is primate-specific. Finally, two DMRs in the *ZNF331* locus were also discovered (Chapter 5).

6.2. Improving the methodology to discover new imprinted genes.

6.2.1. Choice of screening technologies.

The Sequenom and Illumina platforms were both found to be suitable for our project as results confirmed the imprinting status of several known imprinted genes. In total, 1536 SNPs were assayed with the Illumina array. The analysis of data necessitated some arbitrary cut-offs and this pruning left a total of 261 informative SNPs (214 genes) for subsequent analysis. Recently, in human, a higher-throughput Illumina assay called Infinium, was used to screen LCLs and skin fibroblasts of a CEPH family for imprinting (MORCOS *et al.* 2011). This assay tests one million SNPs and so covers the genome. They found two new imprinted genes in LCLs (*ZDBF2* and *SGK2*) and three new ones in fibroblasts (*NAT15*, *RTL1* and *MEG8*). For three of these, namely *ZDBF2*, *RTL1* and *MEG8,* there was prior evidence for imprinting. Despite covering roughly 50% of the genome per tissue, they found only two completely new imprinted genes *NAT15* and *SGK2* that lie in close vicinity to known IGs (*ZNF597* and *L3MBTL* respectively) (MORCOS *et al.* 2011).

The platforms we used were adapted to test candidates that are sufficiently expressed in the tissue tested. Higher-throughput technologies and deep sequencing are now available to screen the whole genome in one go. As I will discuss in the next paragraph, bioinformatics predictions were perfectible and additional criteria should have been incorporated to increase our chances to discover new imprinted genes.

6.2.2. Predictions and choice of candidates.

6.2.2.1. Luedi's predictions.

6.2.2.1.1. *Murine predictions.*

In total, across the two platforms, 155 human orthologues out of the 600 mouse candidates of Luedi et al. were studied experimentally here. Only one revealed itself to be partially imprinted in the human term placenta.

Luedi et al. predicted 600 genes to be imprinted out of 23,788 murine autosomal annotated genes (LUEDI *et al.* 2005). They based their predictions on an old version of Ensembl (Ensembl Mouse Genome version 16) in which repeats were poorly annotated (P. Flicek, Ensembl team, EBI, personal communication). As the prediction algorithm highlighted the importance of repeats in the detection of imprinted genes, this fact has become a serious concern and may explain the poor yield of our screen. Furthermore, as Luedi and colleagues used an old mouse genome build that is no longer accessible in the Ensembl archives, only ~70.5% of the predicted genes could be retrieved with the more recent Ensembl version (Ensembl m35) that was used in the present study. It is also of concern that some old Ensembl gene identification numbers have been re-used for other genes later on (P. Flicek, Ensembl team, EBI, personal communication). An important obstacle to test a large portion of the Luedi's candidates was also the absence of an exonic (or UTR) SNP with a minor allelic frequency of 15% or more (to be statistically testable on 24 trios) (i.e. an informative SNP).

The lack of specificity could be due to the fact that the method used by Luedi et al. also relies on two *small* sets of genes (known imprinted genes and biallelically expressed genes) to train the classifier. Misclassification of genes in one or the other category could have significantly distorted the algorithm prediction accuracy. Considering that the authors relied on the literature to determine which genes are imprinted and which are not and that, as they pointed out themselves, genes can be imprinted only in one specific tissue, at one developmental stage or that only one of their specific isoforms can be imprinted, we directly see another loophole in their algorithm training strategy.

Considering all these limitations, good bioinformatics predictions of imprinting status based on DNA sequence features seem to be an illusory goal. Maybe in the future, one will retrospectively find the correct predictive algorithm when all imprinted genes will be known and/or when other more relevant features will be added to train the software (e.g., epigenetic signatures).

Out of the 155 candidate genes tested in the present work, one gene, *PHACTR2*, exhibited partial imprinting in the term placenta. In another study, also using Luedi's murine predictions, another gene, *KCNK9*, was found imprinted in the mouse and human brain (RUF *et al.* 2007). These two genes alone provide very limited evidence for the validity of the predictions, how can this be explained?

The short list of 16 candidate genes tested by Ruf et al. was refined according to their presence in Luedi's list, possible biomedical relevance *and* their clustering with other predicted candidates (Table 19). Ruf and colleagues used Quantification of Allele-Specific Expression by Pyrosequencing (QUASEP) of RT-PCR products derived from (C57BL/6 x Cast/Ei) F1, C57BL/6 and Cast/Ei E11.5 whole embryos (RUF *et al.* 2007). They then confirmed the imprinted expression of *Kcnk9* by conventional sequencing of RT-PCR products. Hence, they chose a relatively small number of candidates that had several lines of evidence suggesting imprinting and found only one of these 16 genes to be imprinted.

Table 20: List of candidate genes prioritised by Ruf et al.
For 16 out of their 18 selected genes, Ruf et al. found expressed SNPs. If the gene has been tested in the present study, the results obtained are listed in the last column.

Mouse Gene	Allelic expression	Orthologue tested in present study
Kcnk9	Monoallelic maternal	Not studied
Rarres1	Preferential strain-specific	Intensity < 11.25 on Illumina
Disc1	No expressed SNP	Significant for ASE (FDR bound *p*-value, 0.022) on Sequenom see below
Hes5	No expressed SNP	Not studied
Nkx6-2	Biallelic	Not studied
Ntng2	Biallelic	Not studied
Camk2b	Biallelic	Not studied
Irx4	Biallelic	Not studied
Foxg1	Biallelic	Not studied
Dok1	Biallelic	Not studied
Wnt7b	Biallelic	Not studied
Ly6d	Biallelic	Intensity < 11.25 on Illumina
Gdnf	Biallelic	Not studied
Cdk6	Biallelic	Not studied
Gad2	Biallelic	Not studied
Nppc	Biallelic	Not studied
Prdm16	Biallelic	Intensity < 11.25 on Illumina
Stk32c	Biallelic	Not studied

We note that *DISC1* (Disrupted in Schizophrenia 1) was significant for ASE on the Sequenom platform. It is located on hChr1q42.1 and is a susceptibility gene for psychiatric illness (CHUBB *et al.* 2008). It does not lie in the vicinity of an imprinted gene but *Disc1* was shown to interact *in trans* with known imprinted domains by circular chromosome conformation capture (4C) (ZHAO *et al.* 2006) and it has been implicated in psychiatric disorders. For this last reason, Hayesmoore et al. tested 148 brain samples for imprinting. They had 65 informative samples and showed *DISC1* was not imprinted in human brain

(HAYESMOORE *et al.* 2008). As *DISC1* was present on the Illumina array and had low expression levels, the result obtained on Sequenom was considered to be a false positive.

Similarly, *PHACTR2* is adjacent to *PLAGL1* a known imprinted gene (alias *ZAC*) on hChr6q24.2. Combined with prior observations that imprinted genes often occur in clusters, the data obtained by Ruf et al. and in the present study suggest that, if there are more imprinted genes to be found, they may lie close to other imprinted genes rather than being located in new loci.

Shortlisting our candidates by biomedical relevance or genome localisation would have significantly reduced our list, which was in contradiction with our high-throughput testing strategy.

6.2.2.1.2. *Luedi's human predictions.*

After the design of our Illumina array, Luedi and colleagues generated a human set of predictions with two algorithms trained on human imprinted and biallelic genes DNA sequence sets (LUEDI *et al.* 2007). One algorithm was the same Support Vector Machine algorithm as for the mouse predictions, the other was using Sparse Logistic Regression. One hundred and fifty six candidate genes were predicted to be imprinted by both algorithms. Using high algorithmic scores as a more selective criterion plus a few other criteria (either having also been predicted in mouse, or being on a human chromosome without any known imprinted gene, or being close to a chromosomal marker linked to some 'subjectively selected' disease), they had a high-priority list of five genes. They report having tested only two genes in their list for imprinting: *DLGAP2* and *KCNK9*. They found imprinting for *DLGAP2* in testis and *KCNK9* in the brain (as Ruf et al. in the mouse, see above) (LUEDI *et al.* 2007). *DLGAP2* imprinting has not been tested in human placenta or rodents. It is a post-synaptic density protein of unknown role. Why it is expressed and imprinted in testis is not clear. So with their 156 human 'double algorithm' predictions, they found two new human imprinted genes, which is not different from chance (Fisher exact test, *p* value = 0.05). *KCNK9* was imprinted in both human and mouse. The

dysregulation of *KCNK9* was recently found responsible for Birk Barel mental retardation dysmorphism syndrome (BAREL *et al.* 2008). The disease is caused by mutations in the maternal copy of the gene. The identification of this new imprinted gene has enabled the elucidation of this rare syndrome.

Very recently, Jirtle's group showed that another predicted human gene *FAM50B* in their list of 156 candidates predicted by the two-algorithms strategy is also an imprinted human gene (ZHANG *et al.* 2011). *FAM50B* is located at 6p25.2 and is imprinted in brain, liver, placenta, adrenal, heart, testis (and biallelically expressed in the ovary). This gene was not listed as a candidate imprinted gene in mouse and has not been tested in our study. Three confirmed imprinted genes out of the 156 candidates is now a better result than chance alone (Fisher exact test, p value 0.005). This suggests that Luedi has refined the specificity of his predictions. The Ensembl assembly (Version 20) they used for their human predictions was also probably of much better quality than the mouse Ensembl version (16.30) they used for their mouse predictions.

6.2.2.1.3. *Comparison of Luedi's human and mouse predictions.*

Given that almost all genes that are imprinted in human are also imprinted in mouse (the only confirmed exception to this rule is *L3MBTL*), it is surprising that the mouse and human Luedi's predictions lists overlap each other by only 13 genes. This very small overlap suggests that both prediction sets have a weak positive predictive value. There might be explanations for this discrepancy. First, one set could be better than the other. The human one being more recent and based on two different prediction algorithms would win our favour. One could also argue that the predictions pinpoint species-specific imprinted genes. This is theoretically possible. However, to find numerous species-specific imprinted genes would be quite a revolution in the field. Non-coding features like repeats were used to predict candidates in the mouse and human predictions of Luedi et al. and it is, as stated in the previous paragraph, possible that there were differences in the assembly quality of these features in the versions of the human (Ensembl version 20) and mouse (Ensembl version 16) genomes used for these studies.

In conclusion, we would have first tested the human set of candidates had it been available. For the completeness of Luedi's work, it would have been interesting to test the efficiency of the two algorithms on the more recent assemblies of both genomes.

6.2.2.2. Seoighe's predictions.

Twenty-eight candidates out of the 55 human candidates identified by Seoighe et al. by mining EST databases were tested in the present study in the human term placenta but none was found imprinted. The EST predictions of Seoighe et al. were *a priori* solid (as based on biases of SNPs representations in EST libraries) but *a posteriori* might simply represent EST cloning biases (SEOIGHE *et al.* 2006).

6.2.2.3. Birth weight related genes.

We also incorporated 35 genes that were differentially expressed in the placenta according to the infant birth weight (SOOD *et al.* 2006) but none of these were found to be imprinted.

6.2.2.4. Improving bioinformatics predictions.

Only one of the 183 candidates predicted by bioinformatics methods that we tested was found partially imprinted in placenta. The poor specificity of the bioinformatics predictions in placenta raises two possibilities: either the bioinformatics predictions have low specificity overall or the predictions are identifying imprinting in tissues other than placenta.

The slightly better human predictions and the recent discovery of new imprinted genes in other tissues than placenta (PANT *et al.* 2006; LUEDI *et al.* 2007; RUF *et al.* 2007) suggest that both hypotheses might be correct.

Recently, Soloway's group published predictions that were based on DNA sequence but also on new data available for epigenetic features (HEINTZMAN *et al.* 2007; WEN *et al.* 2008). They took into account GC content, CpG islands, miRNA clusters but also CTCF binding sites and histone modifications (H3K4me3, H3K36me, H3K9me3, H3K27me3) (BRIDEAU *et al.* 2010). They tested a subset of their best candidates in mouse placenta of F1 crosses between polymorphic strains. Three genes were fully imprinted (*Cntn3*, *Scin* and *Th*) and six were partially imprinted with the predominant allele being expressed between 62 and 95% (BRIDEAU *et al.* 2010). This approach yielded a far better success rate than the methods based on sequence features alone.

6.2.3. Improving the choice of tissues to screen for new imprinted genes.

6.2.3.1. Placenta.

Ian Morison predicted that only a few new placental imprinted genes were to be found (MORISON *et al.* 2005). His principal argument was that all human genetic disorders that could fit an imprinted mode of inheritance were explained (at least the culprit imprinted gene had been found). However, recently, new imprinted genes have been identified in the human placenta thanks to allelic assays on arrays: *ZNF331* and *PHACTR2* in this study, *NLRP2* and *OSBL1A* in another study (BJORNSSON *et al.* 2008), and *FAM50B* in placenta and other tissues (ZHANG *et al.* 2011). New imprinted genes have also been found in the mouse placenta (SCHULZ *et al.* 2006; MONK *et al.* 2008; BRIDEAU *et al.* 2010) or in the mouse embryo (BABAK *et al.* 2008).

For placental tissue, study during the first trimester of pregnancy would be a logical next step. Indeed, some genes are imprinted only at the beginning of the pregnancy (MONK *et al.* 2006). Humans usually have only one baby per pregnancy and it has been hypothesised that they might have lost their selective pressure for imprinting, most likely during the third trimester and the lactation period (MONK *et al.* 2006). Furthermore in humans placental invasion happens

between the 8th and 16th week and that period might be the best developmental stage to look for placental imprinted genes, as efficient invasion will enable the placenta to extract more resources from the mother.

New placental imprinted genes are still being found. Another avenue would be to evaluate placental cell-type specific imprinting (GIMELBRANT *et al.* 2007; VAN DIJK *et al.* 2010). Recently, Barbaux et al. have used the same genotyping arrays as Gimelbrant et al. to compare genotypes of cDNA vs.gDNA on human placentas using 250K Affymetrix arrays that contain a very large proportion of intronic SNPs (BARBAUX *et al.* 2012). After filtration of their data only 12% of the SNPs tested could be analysed. This study identified several new imprinted genes: *ZFAT*, *ZFAT-AS1*, *LIN28B*, *ZNF597*, *GLIS3*, *ZC3H12C*, *NTM*, *MAGI2* (BARBAUX *et al.* 2012). It is noteworthy that among these genes, two have a "less stringent" monoallelic expression. For *MAGI2* and *NTM*, only half of the heterozygous samples tested show monoallelic expression.

6.2.3.2. Other tissues.

Luedi et al. have confirmed imprinting for two of their human candidates in testis and brain tissue (LUEDI *et al.* 2007). All isoforms of *DLGAP2* were paternally expressed in the testis of six human conceptuses (63-105 days) and *KCNK9* was confirmed imprinted and maternally expressed in nine informative fetal brain samples (63-98 days conceptuses) (LUEDI *et al.* 2007; RUF *et al.* 2007). Kelly Frazer's group has found three new imprinted genes in lymphocytes (PANT *et al.* 2006). This means that it is possible that mainly *tissue specific* imprinting remains to be found.

Three groups have used RNA-Seq on mouse brain samples. Babak et al. have discovered six novel IG with this genome-wide technology. Most of these were non-coding RNAs in known imprinted loci. Wang et al. have used the same strategy and have identified three novel IGs in known loci. Recently, Gregg et al. have used RNA-Seq to study several mouse brain regions. They have suggested that around 800 new protein-coding and around 400 new non-coding RNA genes could exhibit parent-of-origin allelic effects (GREGG *et al.* 2010).

The interpretation of results, depth of sequencing, technical progress and different statistical approaches explain these discrepant results. Babak et al. have a conservative vision of imprinting. This led them to discard partially imprinted candidate genes. They also didn't envisage more complex parent-of-origin effects (BABAK *et al.* 2008; DEVEALE *et al.* 2012). Wang et al. have only studied whole brains and had a lower sequencing depth (WANG *et al.* 2008b). Gregg at al. had the most comprehensive approach. They studied samples of whole brain (E15), male and female prefrontal cortex and hypothalamus. They have studied imprinting quantitatively. This allowed them to predict many partially imprinted genes. They also studied complex parent-of-origin effects as in Cheverud et al. (e.g., sex-specific effects, polar overdominance)(CHEVERUD *et al.* 2008). At several loci, they confirmed their findings by a second technique (qPCR). Hence, they have studied quantitatively tissue-specific and isoform-specific imprinting in mouse brains and have predicted hundreds of candidates. Even if their results are bound to contain false positives, this is in my view a good screening approach to discover new allelic effects. However, these RNA-Seq-'predictions' should be confirmed by extensive functional studies using a second sensitive and quantitative method.

As discussed above, it is only the study of each tissue at several developmental stages that will reveal the real extent of imprinting.

6.3. Characterisation of parent-of-origin allele specific expression in human placenta.

Some characteristics of the known imprinted genes tested are interesting. The quantitative ASE results for the imprinted genes on the array showed that 'silencing' of the repressed allele was not absolute. It was rather a continuum from complete silencing (e.g. *PEG3*, *H19*, and *MEST*) to partial silencing (e.g., *DLK1*, *IGF2AS*, and *PHACTR2*). These results agree with the work of Lambertini et al. who have seen some expression of the silenced allele for nine imprinted genes including *DLK1* and *H19* in human placenta (LAMBERTINI *et al.* 2008). In fact not much is known about what constitutes the 'normal' spectrum of expression of the 'silent' allele in healthy tissue samples. For example, significant expression of the 'silent' allele of *IGF2* has been previously observed in peripheral blood leucocytes of 10.5% (4/38) of healthy Japanese individuals (SAKATANI *et al.* 2001).

This is a new paradigm for imprinting. Parent-specific allelic expression is a continuum from complete silencing of one parental allele to a parentally biased expression of the two alleles. So absolute imprinting is at the end of the normal distribution of parent-of-origin ASE. The phrase *partially imprinted* has not been used until recently. In the past, results that didn't demonstrate the purest form of imprinting have probably been discarded or labelled as *biallelic*. This new pattern questions the definition of imprinting. When is it absolute imprinting and when is it only partial imprinting? It is very likely that some partially imprinted genes could still be found.

6.4. Relevance of genes exhibiting ASE in placenta.

6.4.1. Physiological relevance of ASE in placenta.

Using moderated *t*-statistics, thirty-nine genes were statistically significant for ASE in placenta in this study. These results suggest that our 39 ASE genes are likely to be important for placental physiology and support further study of their role in normal and pathological pregnancy. For example, further examination of the haplotypes, ASE, and expression levels of these 39 genes in relation to fetal growth could be of interest. Without any functional data on each of these genes, it is impossible to know whether they interact and which ones might be key candidate genes for phenotypic variability.

6.4.2. Relevance of bipolar ASE.

We analysed five modes of ASE (imprinted, partial imprinting, preferential, monoallelic random, random ASE). Recently, Cheverud and colleagues suggested that different bipolar modes of ASE could exist (CHEVERUD *et al.* 2008; WOLF *et al.* 2008; LAWSON *et al.* 2011). Bipolar ASE shows allele specific bias depending first on the parent-of-origin of the allele and second on heterozygous or homozygous status for this allele (a mode of allelic expression inheritance that was previously only known in the callipyge sheep (COCKETT *et al.* 1996)). Considering the bipolar associated growth and metabolic phenotypes described by Cheverud et al. in the adult mouse (CHEVERUD *et al.* 2008), it will be interesting to explore bipolar ASE in human tissues. However, the platforms used in this study would need to test many more trios with more replicates to approach the precision required to investigate such complex ASE patterns.

6.4.3. Relevance of monoallelic random ASE.

We found eight genes randomly biased between individuals. For six genes, the biases of expression were small (60-40 to 80-20). This represents roughly 1.9%

of the genes tested. Gimelbrant et al. using hybridisation of nuclear RNA of lymphoblasts on Affymetrix arrays (both exonic and intronic SNPs), reported that 10% of autosomal genes presented random monoallelic expression clonally stable in descendants (GIMELBRANT *et al.* 2007). The Frazer group had found similar variability between CEPH lymphoblast clones too (PANT *et al.* 2006; POLLARD *et al.* 2008). We could interpret their result more conservatively and say that 10% of autosomal genes present epigenetic instability in lymphoblast clones. Gimelbrant's *in vivo* results on 1mm3 placental tissue seem more interesting. RNA FISH experiments suggest a highly variable level of allelic expression in the cells of a same tissue (JOUVENOT *et al.* 1999; OSBORNE *et al.* 2004). The placental studies of Gimelbrant et al. confirm this point and suggest that each tissue is made of heterogeneous cells not expressing the same allelic ratio as their neighbours (GIMELBRANT *et al.* 2007). However, this is not really different than the random mosaicism for X inactivation, and biologically what is probably relevant for fetal growth is the *general ratio* of allelic expression in the whole placenta (or at least broader regions than 1 mm^3). Only highly skewed whole placental expression would be medically relevant as seen for the skewing of X inactivation. Furthermore, Gimelbrant et al. studied *nuclear* RNA that comprises all non-coding, antisense and intergenic transcription, so that the transcription could be potentially much more unstable (GIMELBRANT *et al.* 2007).

In conclusion, using 100 µg of tissue for the extraction of total RNA, we have found much less random monoallelic expression for protein coding genes than claimed by others on cell lines.

The results of our study and all the ones cited here above suggest that only strong allelic effect such as imprinting, tissue specific promoters or allelic preferential expression might be a medically relevant mode of ASE.

6.4.4. Normal distribution of ASE.

Biological processes are normally distributed and the data obtained fits this distribution (Figure 69). As ASE observations accumulate, the allelic ratio distribution seems to be Gaussian (with different axes: haplotypic ratio, parental ratio). High-throughput quantitative ASE information could bring a re-definition of types of ASE. For example, in the future, a threshold for calling a gene "*absolutely"* imprinted could be defined. In addition, biological and medical relevance of variegating ASE ratios could be studied.

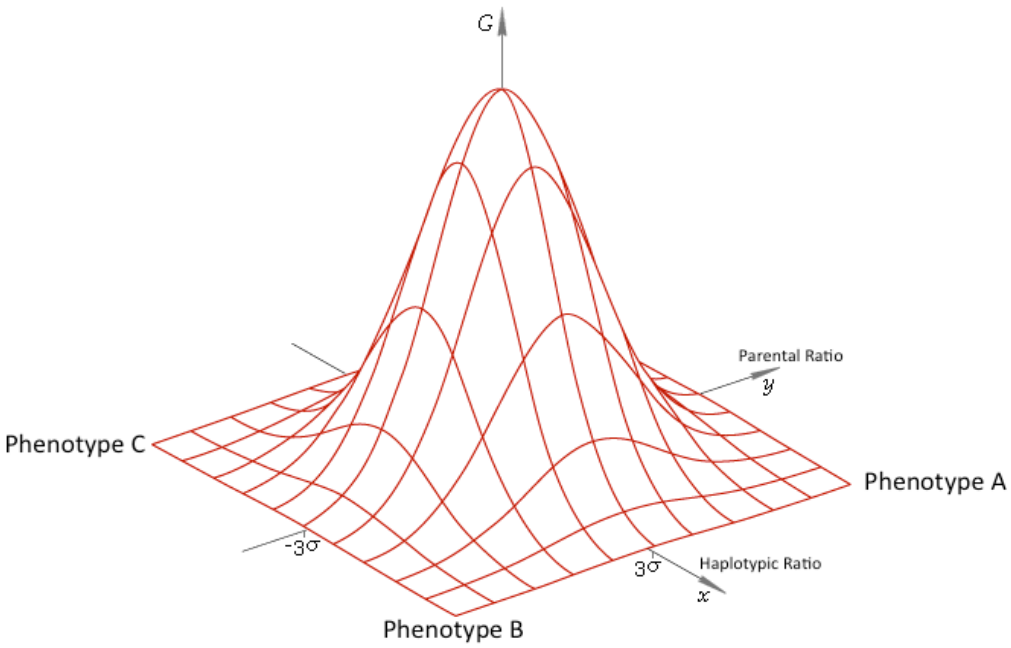

Figure 64: Normal distribution hypothesis.

The x-axis represents the haplotypic ratio, the y-axis the parental ratio. Picture adapted from librow.com.

6.5. Conclusion and future work.

We have found that allelic imbalances in expression are common in the candidates we analysed in the human term placenta and that true monoallelic expression (imprinted or random) is a rare phenomenon. We studied experimentally 183 genes identified as candidates for imprinted expression by prior bioinformatics approaches (LUEDI *et al.* 2005; SEOIGHE *et al.* 2006). We have found only two new (partially) imprinted genes in the human placenta (1%), while ASE was present in 18% of the candidate genes passing our quality control criteria. Such levels of ASE are similar to the results seen in cell lines or other somatic tissues (LO *et al.* 2003; PANT *et al.* 2006; GIMELBRANT *et al.* 2007; BJORNSSON *et al.* 2008; POLLARD *et al.* 2008).

In future screens for imprinted genes, it could be interesting to study:

- genes predicted to be imprinted by two algorithms on the most recent human genome assemblies.
- genes located close to known imprinted genes
- genes sharing the same epigenetic features as known imprinted genes
- genes predicted to be imprinted in both human and mouse
- genes linked to a fetal growth or imprinting network (VARRAULT *et al.* 2006; ZHAO *et al.* 2006)
- genes or gene loci that have recently evolved in human (cfr the *ZNF331-C19MC* primate-specific locus).
- all genes using a high throughput system, in all tissues available.

However, thanks to the recent advances in high-throughput sequencing, genome-wide surveys of imprinting are now being undertaken by-passing the need for candidate genes bioinformatics predictions (BABAK *et al.* 2008; NICA *et al.* 2011). The deep sequencing approach will enable a thorough characterisation of imprinting in every tissue for every transcript. One could even dream about "single cell" ASE screening methods.

In conclusion, pushed by the medical importance of imprinted genes, we embarked for a high-throughput screen of imprinted candidate genes based on bioinformatics predictions. While our functional approach worked, it seems that the different bioinformatics predictions we used to obtain candidate genes have been over-predictive at best. Hence, agnostic and more high-throughput functional screens need to be undertaken to comprehensively establish the extent of human imprinting.

References

Limma. http://bioinf.wehi.edu.au/limma.

R Foundation for Statistical Computing, Vienna, Austria R Development Core Team. R: A Language and Environment for Statistical Computing **2009:** http://www.R-project.org.

ALKAN, C., J. M. KIDD, T. MARQUES-BONET, G. AKSAY, F. ANTONACCI *et al.*, 2009 Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet **41:** 1061-1067.

ALLEN, E., S. HORVATH, F. TONG, P. KRAFT, E. SPITERI *et al.*, 2003 High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proc Natl Acad Sci U S A **100:** 9940-9945.

ALLEN, P. B., A. T. GREENFIELD, P. SVENNINGSSON, D. C. HASPESLAGH and P. GREENGARD, 2004 Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins. Proc Natl Acad Sci U S A **101:** 7187-7192.

ALTSHULER, D., and M. DALY, 2007 Guilt beyond a reasonable doubt. Nature genetics **39:** 813-815.

ALTSHULER, D., V. J. POLLARA, C. R. COWLES, W. J. VAN ETTEN, J. BALDWIN *et al.*, 2000 An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature **407:** 513-516.

ALTSHULER, D. M., R. A. GIBBS, L. PELTONEN, E. DERMITZAKIS, S. F. SCHAFFNER *et al.*, 2010 Integrating common and rare genetic variation in diverse human populations. Nature **467:** 52-58.

ANDERSEN, A. N., V. GOOSSENS, L. GIANAROLI, R. FELBERBAUM, J. DE MOUZON *et al.*, 2007 Assisted reproductive technology in Europe, 2003. Results generated from European registers by ESHRE. Hum Reprod **22:** 1513-1525.

APOSTOLIDOU, S., S. ABU-AMERO, K. O'DONOGHUE, J. FROST, O. OLAFSDOTTIR *et al.*, 2007 Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. Journal of Molecular Medicine (Berlin, Germany) **85:** 379-387.

ARIMA, T., R. A. DREWELL, K. L. ARNEY, J. INOUE, Y. MAKITA *et al.*, 2001 A conserved imprinting control region at the HYMAI/ZAC domain is implicated in transient neonatal diabetes mellitus. Hum Mol Genet **10:** 1475-1483.

ARIMA, T., R. A. DREWELL, M. OSHIMURA, N. WAKE and M. A. SURANI, 2000 A novel imprinted gene, HYMAI, is located within an imprinted domain on human chromosome 6 containing ZAC. Genomics **67:** 248-255.

ARIMA, T., K. YAMASAKI, R. M. JOHN, K. KATO, K. SAKUMI *et al.*, 2006 The human HYMAI/PLAGL1 differentially methylated region acts as an imprint control region in mice. Genomics **88:** 650-658.

ARNAUD, P., K. HATA, M. KANEDA, E. LI, H. SASAKI *et al.*, 2006 Stochastic imprinting in the progeny of Dnmt3L-/- females. Hum Mol Genet **15:** 589-598.

BABAK, T., B. DEVEALE, C. ARMOUR, C. RAYMOND, M. A. CLEARY *et al.*, 2008 Global survey of genomic imprinting by transcriptome sequencing. Curr Biol **18:** 1735-1741.

BARBAUX, S., G. GASCOIN-LACHAMBRE, C. BUFFAT, P. MONNIER, F. MONDON *et al.*, 2012 A genome-wide approach reveals novel imprinted genes expressed in the human placenta. Epigenetics **7:** 1079-1090.

BAREL, O., S. A. SHALEV, R. OFIR, A. COHEN, J. ZLOTOGORA *et al.*, 2008 Maternally inherited Birk Barel mental retardation dysmorphism syndrome caused by a mutation in the genomically imprinted potassium channel KCNK9. Am J Hum Genet **83:** 193-199.

BARKER, D. J., 2006 Adult consequences of fetal growth restriction. Clin Obstet Gynecol **49:** 270-283.

BARKER, D. J., J. G. ERIKSSON, T. FORSEN and C. OSMOND, 2002 Fetal origins of adult disease: strength of effects and biological basis. Int J Epidemiol **31:** 1235-1239.

BARKER, D. J., P. D. WINTER, C. OSMOND, B. MARGETTS and S. J. SIMMONDS, 1989 Weight in infancy and death from ischaemic heart disease. Lancet **2:** 577-580.

BARLOW, D. P., 1994 Imprinting: a gamete's point of view. Trends Genet **10:** 194-199.

BARLOW, D. P., 1995 Gametic imprinting in mammals. Science **270:** 1610-1613.

BARLOW, D. P., R. STOGER, B. G. HERRMANN, K. SAITO and N. SCHWEIFER, 1991 The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature **349:** 84-87.

BAROUX, C., V. GAGLIARDINI, D. R. PAGE and U. GROSSNIKLAUS, 2006 Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. Genes Dev **20:** 1081-1086.

BARTOLOMEI, M. S., S. ZEMEL and S. M. TILGHMAN, 1991 Parental imprinting of the mouse H19 gene. Nature **351:** 153-155.

BARTON, S. C., M. A. SURANI and M. L. NORRIS, 1984 Role of paternal and maternal genomes in mouse development. Nature **311:** 374-376.

BEAUDET, A. L., and Y. H. JIANG, 2002 A rheostat model for a rapid and reversible form of imprinting-dependent evolution. Am J Hum Genet **70:** 1389-1397.

BELL, A. C., and G. FELSENFELD, 2000 Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature **405:** 482-485.

BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological)**:** 289-300.

BENTWICH, I., A. AVNIEL, Y. KAROV, R. AHARONOV, S. GILAD *et al.*, 2005 Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet **37:** 766-770.

BEREZIKOV, E., F. THUEMMLER, L. W. VAN LAAKE, I. KONDOVA, R. BONTROP *et al.*, 2006 Diversity of microRNAs in human and chimpanzee brain. Nat Genet **38:** 1375-1377.

BESTOR, T. H., 1998 The host defence function of genomic methylation patterns. Novartis Found Symp **214:** 187-195; discussion 195-189, 228-132.

BESTOR, T. H., 2000 The DNA methyltransferases of mammals. Hum Mol Genet **9:** 2395-2402.

BININDA-EMONDS, O. R., M. CARDILLO, K. E. JONES, R. D. MACPHEE, R. M. BECK *et al.*, 2007 The delayed rise of present-day mammals. Nature **446:** 507-512.

BIRNEY, E., J. A. STAMATOYANNOPOULOS, A. DUTTA, R. GUIGO, T. R. GINGERAS *et al.*, 2007 Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447:** 799-816.

BJORNSSON, H. T., T. J. ALBERT, C. M. LADD-ACOSTA, R. D. GREEN, M. A. RONGIONE *et al.*, 2008 SNP-specific array-based allele-specific expression analysis. Genome Res **18:** 771-779.

BOCHUKOVA, E. G., N. HUANG, J. KEOGH, E. HENNING, C. PURMANN *et al.*, 2010 Large, rare chromosomal deletions associated with severe early-onset obesity. Nature **463:** 666-670.

BOOTH, H. A., and P. W. HOLLAND, 2007 Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. Gene **387:** 7-14.

BORCHERT, G. M., W. LANIER and B. L. DAVIDSON, 2006 RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol **13:** 1097-1101.

BORSANI, G., R. TONLORENZI, M. C. SIMMLER, L. DANDOLO, D. ARNAUD *et al.*, 1991 Characterization of a murine gene expressed from the inactive X chromosome. Nature **351:** 325-329.

BORTOLIN-CAVAILLE, M. L., M. DANCE, M. WEBER and J. CAVAILLE, 2009 C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. Nucleic Acids Res **37:** 3464-3473.

BOURC'HIS, D., and T. H. BESTOR, 2006 Origins of extreme sexual dimorphism in genomic imprinting. Cytogenet Genome Res **113:** 36-40.

BOURC'HIS, D., G. L. XU, C. S. LIN, B. BOLLMAN and T. H. BESTOR, 2001 Dnmt3L and the establishment of maternal genomic imprints. Science **294:** 2536-2539.

BOWCOCK, A. M., 2007 Genomics: guilt by association. Nature **447:** 645-646.

BOWDIN, S., C. ALLEN, G. KIRBY, L. BRUETON, M. AFNAN *et al.*, 2007 A survey of assisted reproductive technology births and imprinting disorders. Hum Reprod **22:** 3237-3240.

BRAY, N. J., P. R. BUCKLAND, M. J. OWEN and M. C. O'DONOVAN, 2003 Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet **113:** 149-153.

BRESLAUER, K. J., R. FRANK, H. BLOCKER and L. A. MARKY, 1986 Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A **83:** 3746-3750.

BRIDEAU, C. M., K. E. EILERTSON, J. A. HAGARMAN, C. D. BUSTAMANTE and P. D. SOLOWAY, 2010 Successful computational prediction of novel imprinted genes from epigenomic features. Mol Cell Biol **30:** 3357-3370.

BROCKDORFF, N., A. ASHWORTH, G. F. KAY, P. COOPER, S. SMITH *et al.*, 1991 Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature **351:** 329-331.

BROWN, C. J., A. BALLABIO, J. L. RUPERT, R. G. LAFRENIERE, M. GROMPE *et al.*, 1991 A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature **349:** 38-44.

BUTLER, M. G., W. FISCHER, N. KIBIRYEVA and D. C. BITTEL, 2008 Array comparative genomic hybridization (aCGH) analysis in Prader-Willi syndrome. Am J Med Genet A **146:** 854-860.

CARREL, L., and H. F. WILLARD, 2005 X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature **434:** 400-404.

CARTER, N. P., 2007 Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet **39:** S16-21.

CATTANACH, B. M., and M. KIRK, 1985 Differential activity of maternally and paternally derived chromosome regions in mice. Nature **315:** 496-498.

CEDAR, H., and Y. BERGMAN, 2009 Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet **10:** 295-304.

CHAILLET, J. R., T. F. VOGT, D. R. BEIER and P. LEDER, 1991 Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell **66:** 77-83.

CHEDIN, F., M. R. LIEBER and C. L. HSIEH, 2002 The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A **99:** 16916-16921.

CHEN, T., and E. LI, 2004 Structure and function of eukaryotic DNA methyltransferases. Curr Top Dev Biol **60:** 55-89.

CHESS, A., I. SIMON, H. CEDAR and R. AXEL, 1994 Allelic inactivation regulates olfactory receptor gene expression. Cell **78:** 823-834.

CHEUNG, V. G., R. S. SPIELMAN, K. G. EWENS, T. M. WEBER, M. MORLEY *et al.*, 2005 Mapping determinants of human gene expression by regional and genome-wide association. Nature **437:** 1365-1369.

CHEVERUD, J. M., R. HAGER, C. ROSEMAN, G. FAWCETT, B. WANG *et al.*, 2008 Genomic imprinting effects on adult body composition in mice. Proc Natl Acad Sci U S A **105:** 4253-4258.

CHIANG, D. Y., G. GETZ, D. B. JAFFE, M. J. O'KELLY, X. ZHAO *et al.*, 2009 High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods **6:** 99-103.

CHOI, J. D., L. A. UNDERKOFFLER, A. J. WOOD, J. N. COLLINS, P. T. WILLIAMS *et al.*, 2005 A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island. Mol Cell Biol **25:** 5514-5522.

CHOW, J., and E. HEARD, 2009 X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol **21:** 359-366.

CHUBB, J. E., N. J. BRADSHAW, D. C. SOARES, D. J. PORTEOUS and J. K. MILLAR, 2008 The DISC locus in psychiatric illness. Mol Psychiatry **13:** 36-64.

CICCONE, D. N., and T. CHEN, 2009 Histone lysine methylation in genomic imprinting. Epigenetics **4:** 216-220.

CLEMSON, C. M., J. A. MCNEIL, H. F. WILLARD and J. B. LAWRENCE, 1996 XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol **132:** 259-275.

COAN, P. M., G. J. BURTON and A. C. FERGUSON-SMITH, 2005 Imprinted genes in the placenta--a review. Placenta **26 Suppl A:** S10-20.

COCKETT, N. E., S. P. JACKSON, T. L. SHAY, F. FARNIR, S. BERGHMANS *et al.*, 1996 Polar overdominance at the ovine callipyge locus. Science **273:** 236-238.

CONSORTIUM, E. P., E. BIRNEY, J. A. STAMATOYANNOPOULOS, A. DUTTA, R. GUIGO *et al.*, 2007 Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447:** 799-816.

COX, G. F., J. BURGER, V. LIP, U. A. MAU, K. SPERLING *et al.*, 2002 Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet **71:** 162-164.

CROUSE, H. V., 1960 The Controlling Element in Sex Chromosome Behavior in Sciara. Genetics **45:** 1429-1443.

CUI, H., I. L. HORON, R. OHLSSON, S. R. HAMILTON and A. P. FEINBERG, 1998 Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability. Nat Med **4:** 1276-1280.

CUSCO, I., R. COROMINAS, M. BAYES, R. FLORES, N. RIVERA-BRUGUES *et al.*, 2008 Copy number variation at the 7q11.23 segmental duplications is a susceptibility factor for the Williams-Beuren syndrome deletion. Genome Res **18:** 683-694.

DAVIES, W., A. ISLES, R. SMITH, D. KARUNADASA, D. BURRMANN *et al.*, 2005 Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. Nature genetics **37:** 625-629.

DAVIS, T. L., G. J. YANG, J. R. MCCARREY and M. S. BARTOLOMEI, 2000 The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. Hum Mol Genet **9:** 2885-2894.

DEAN, W. K., G; REIK W, 2001 *Genomic imprinting: methods and protocols*. The humana Press, Inc., NJ.

DEBAUN, M. R., E. L. NIEMITZ and A. P. FEINBERG, 2003 Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. Am J Hum Genet **72:** 156-160.

DECHIARA, T. M., E. J. ROBERTSON and A. EFSTRATIADIS, 1991 Parental imprinting of the mouse insulin-like growth factor II gene. Cell **64:** 849-859.

DEVEALE, B., D. VAN DER KOOY and T. BABAK, 2012 Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet **8:** e1002600.

DIATLOFF-ZITO, C., A. NICOLE, G. MARCELIN, H. LABIT, E. MARQUIS *et al.*, 2007 Genetic and epigenetic defects at the 6q24 imprinted locus in a cohort of 13 patients with transient neonatal diabetes: new hypothesis raised by the finding of a unique case with hemizygotic deletion in the critical region. J Med Genet **44:** 31-37.

DIMAS, A. S., B. E. STRANGER, C. BEAZLEY, R. D. FINN, C. E. INGLE *et al.*, 2008 Modifier effects between regulatory and protein-coding variation. PLoS Genet **4:** e1000244.

DIPLAS, A. I., L. LAMBERTINI, M. J. LEE, R. SPERLING, Y. L. LEE *et al.*, 2009 Differential expression of imprinted genes in normal and IUGR human placentas. Epigenetics **4**.

DIXON, A. L., L. LIANG, M. F. MOFFATT, W. CHEN, S. HEATH *et al.*, 2007 A genome-wide association study of global gene expression. Nat Genet **39:** 1202-1207.

DJURIC, U., O. EL-MAARRI, B. LAMB, R. KUICK, M. SEOUD *et al.*, 2006 Familial molar tissues due to mutations in the inflammatory gene, NALP7, have normal postzygotic DNA methylation. Hum Genet **120:** 390-395.

DOHERTY, A. S., M. R. MANN, K. D. TREMBLAY, M. S. BARTOLOMEI and R. M. SCHULTZ, 2000 Differential effects of culture on imprinted H19 expression in the preimplantation mouse embryo. Biol Reprod **62:** 1526-1535.

DOORNBOS, M. E., S. M. MAAS, J. MCDONNELL, J. P. VERMEIDEN and R. C. HENNEKAM, 2007 Infertility, assisted reproduction technologies and imprinting disturbances: a Dutch study. Hum Reprod **22:** 2476-2480.

DUNNING, M. J., M. L. SMITH, M. E. RITCHIE and S. TAVARE, 2007 beadarray: R classes and methods for Illumina bead-based data. Bioinformatics **23:** 2183-2184.

DUNSON, D. B., D. D. BAIRD and B. COLOMBO, 2004 Increased infertility with age in men and women. Obstet Gynecol **103:** 51-56.

EMILSSON, V., G. THORLEIFSSON, B. ZHANG, A. S. LEONARDSON, F. ZINK *et al.*, 2008 Genetics of gene expression and its effect on disease. Nature **452:** 423-428.

ENGEL, N., J. L. THORVALDSEN and M. S. BARTOLOMEI, 2006 CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. Hum Mol Genet **15:** 2945-2954.

ENSENAUER, R. E., A. ADEYINKA, H. C. FLYNN, V. V. MICHELS, N. M. LINDOR *et al.*, 2003 Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. Am J Hum Genet **73:** 1027-1040.

ESUMI, S., N. KAKAZU, Y. TAGUCHI, T. HIRAYAMA, A. SASAKI *et al.*, 2005 Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. Nat Genet **37:** 171-176.

FAN, J. B., K. L. GUNDERSON, M. BIBIKOVA, J. M. YEAKLEY, J. CHEN *et al.*, 2006 Illumina universal bead arrays. Methods Enzymol **410:** 57-73.

FAN, J. B., A. OLIPHANT, R. SHEN, B. G. KERMANI, F. GARCIA *et al.*, 2003 Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol **68:** 69-78.

FAN, J. B., J. M. YEAKLEY, M. BIBIKOVA, E. CHUDIN, E. WICKHAM *et al.*, 2004 A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome research **14:** 878-885.

FEDORIW, A. M., P. STEIN, P. SVOBODA, R. M. SCHULTZ and M. S. BARTOLOMEI, 2004 Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. Science **303:** 238-240.

FEINBERG, A. P., 2007 An epigenetic approach to cancer etiology. Cancer J **13:** 70-74.

FELDMANN, J., A. M. PRIEUR, P. QUARTIER, P. BERQUIN, S. CERTAIN *et al.*, 2002 Chronic infantile neurological cutaneous and articular syndrome is caused by mutations in CIAS1, a gene highly expressed in polymorphonuclear cells and chondrocytes. Am J Hum Genet **71:** 198-203.

FERGUSON-SMITH, A. C., 2011 Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet **12:** 565-575.

FERGUSON-SMITH, A. C., B. M. CATTANACH, S. C. BARTON, C. V. BEECHEY and M. A. SURANI, 1991 Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature **351:** 667-670.

FERGUSON-SMITH, A. C., H. SASAKI, B. M. CATTANACH and M. A. SURANI, 1993 Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature **362:** 751-755.

FORTIER, A. L., F. L. LOPES, N. DARRICARRERE, J. MARTEL and J. M. TRASLER, 2008 Superovulation alters the expression of imprinted genes in the midgestation mouse placenta. Hum Mol Genet **17:** 1653-1665.

FRAGA, M. F., E. BALLESTAR, M. F. PAZ, S. ROPERO, F. SETIEN *et al.*, 2005 Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A **102:** 10604-10609.

FRAZER, K. A., D. G. BALLINGER, D. R. COX, D. A. HINDS, L. L. STUVE *et al.*, 2007 A second generation human haplotype map of over 3.1 million SNPs. Nature **449:** 851-861.

FROMMER, M., L. E. MCDONALD, D. S. MILLAR, C. M. COLLIS, F. WATT *et al.*, 1992 A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A **89:** 1827-1831.

FROST, J. M., and G. E. MOORE, 2010 The importance of imprinting in the human placenta. PLoS Genet **6:** e1001015.

GARDNER, R. J., D. J. MACKAY, A. J. MUNGALL, C. POLYCHRONAKOS, R. SIEBERT *et al.*, 2000 An imprinted locus associated with transient neonatal diabetes mellitus. Hum Mol Genet **9:** 589-596.

GE, B., S. GURD, T. GAUDIN, C. DORE, P. LEPAGE *et al.*, 2005 Survey of allelic expression using EST mining. Genome Res **15:** 1584-1591.

GEORGIADES, P., M. WATKINS, G. J. BURTON and A. C. FERGUSON-SMITH, 2001 Roles for genomic imprinting and the zygotic genome in placental development. Proc Natl Acad Sci U S A **98:** 4522-4527.

GICQUEL, C., V. GASTON, J. MANDELBAUM, J. P. SIFFROI, A. FLAHAULT *et al.*, 2003 In vitro fertilization may increase the risk of Beckwith-Wiedemann syndrome related to the abnormal imprinting of the KCN1OT gene. Am J Hum Genet **72:** 1338-1341.

GILES, K. E., H. GOWHER, R. GHIRLANDO, C. JIN and G. FELSENFELD, 2010 Chromatin Boundaries, Insulators, and Long-Range Interactions in the Nucleus. Cold Spring Harb Symp Quant Biol.

GIMELBRANT, A., J. N. HUTCHINSON, B. R. THOMPSON and A. CHESS, 2007 Widespread monoallelic expression on human autosomes. Science **318:** 1136-1140.

GOLDSTEIN, B. A., A. E. HUBBARD, A. CUTLER and L. F. BARCELLOS, 2010 An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. BMC Genet **11:** 49.

GONZALEZ, E., H. KULKARNI, H. BOLIVAR, A. MANGANO, R. SANCHEZ *et al.*, 2005 The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science **307:** 1434-1440.

GORING, H. H., J. E. CURRAN, M. P. JOHNSON, T. D. DYER, J. CHARLESWORTH *et al.*, 2007 Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet **39:** 1208-1216.

GOSDEN, R., J. TRASLER, D. LUCIFERO and M. FADDY, 2003 Rare congenital disorders, imprinted genes, and assisted reproductive technology. Lancet **361:** 1975-1977.

GREALLY, J. M., 2002 Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. Proceedings of the National Academy of Sciences of the United States of America **99:** 327-332.

GREGG, C., J. ZHANG, B. WEISSBOURD, S. LUO, G. P. SCHROTH *et al.*, 2010 High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science **329:** 643-648.

GRUMMT, I., 2007 Different epigenetic layers engage in complex crosstalk to define the epigenetic state of mammalian rRNA genes. Hum Mol Genet **16 Spec No 1:** R21-27.

GU, T. P., F. GUO, H. YANG, H. P. WU, G. F. XU *et al.*, 2011 The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature **477:** 606-610.

GUNDERSON, K. L., F. J. STEEMERS, G. LEE, L. G. MENDOZA and M. S. CHEE, 2005 A genome-wide scalable SNP genotyping assay using microarray technology. Nature genetics **37:** 549-554.

HAIG, D., 2004 Genomic imprinting and kinship: how good is the evidence? Annu Rev Genet **38:** 553-585.

HAJKOVA, P., K. ANCELIN, T. WALDMANN, N. LACOSTE, U. C. LANGE *et al.*, 2008 Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature **452:** 877-881.

HAJKOVA, P., S. ERHARDT, N. LANE, T. HAAF, O. EL-MAARRI *et al.*, 2002 Epigenetic reprogramming in mouse primordial germ cells. Mech Dev **117:** 15-23.

HARK, A. T., C. J. SCHOENHERR, D. J. KATZ, R. S. INGRAM, J. M. LEVORSE *et al.*, 2000 CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature **405:** 486-489.

HARRIS, R. A., T. WANG, C. COARFA, R. P. NAGARAJAN, C. HONG *et al.*, 2010 Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol **28:** 1097-1105.

HATA, K., M. OKANO, H. LEI and E. LI, 2002 Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development **129:** 1983-1993.

HAYASHIZAKI, Y., S. HIROTSUNE, Y. OKAZAKI, H. SHIBATA, A. AKASAKO *et al.*, 1994 A genetic linkage map of the mouse using restriction landmark genomic scanning (RLGS). Genetics **138:** 1207-1238.

HAYESMOORE, J. B., N. J. BRAY, M. J. OWEN and M. C. O'DONOVAN, 2008 DISC1 mRNA expression is not influenced by common Cis-acting regulatory polymorphisms or imprinting. Am J Med Genet B Neuropsychiatr Genet **147B:** 1065-1069.

HAYWARD, B. E., M. KAMIYA, L. STRAIN, V. MORAN, R. CAMPBELL *et al.*, 1998 The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. Proc Natl Acad Sci U S A **95:** 10038-10043.

HEINTZMAN, N. D., R. K. STUART, G. HON, Y. FU, C. W. CHING *et al.*, 2007 Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet **39:** 311-318.

HENCKEL, A., and P. ARNAUD, 2010 Genome-wide identification of new imprinted genes. Brief Funct Genomics **9:** 304-314.

HENRY, I., C. BONAITI-PELLIE, V. CHEHENSSE, C. BELDJORD, C. SCHWARTZ *et al.*, 1991 Uniparental paternal disomy in a genetic cancer-predisposing syndrome. Nature **351:** 665-667.

HENRY, I., M. JEANPIERRE, P. COUILLIN, F. BARICHARD, J. L. SERRE *et al.*, 1989 Molecular definition of the 11p15.5 region involved in Beckwith-Wiedemann syndrome and probably in predisposition to adrenocortical carcinoma. Hum Genet **81:** 273-277.

HERZING, L. B., S. J. KIM, E. H. COOK, JR. and D. H. LEDBETTER, 2001 The human aminophospholipid-transporting ATPase gene ATP10C maps

adjacent to UBE3A and exhibits similar imprinted expression. Am J Hum Genet **68:** 1501-1505.

HERZING, L. B., J. T. ROMER, J. M. HORN and A. ASHWORTH, 1997 Xist has properties of the X-chromosome inactivation centre. Nature **386:** 272-275.

HOLLANDER, G. A., S. ZUKLYS, C. MOREL, E. MIZOGUCHI, K. MOBISSON *et al.*, 1998 Monoallelic expression of the interleukin-2 locus. Science **279:** 2118-2121.

HOLMES, R., Y. CHANG and P. D. SOLOWAY, 2006 Timing and sequence requirements defined for embryonic maintenance of imprinted DNA methylation at Rasgrf1. Mol Cell Biol **26:** 9564-9570.

HOLMQUIST, G. P., 1987 Role of replication time in the control of tissue-specific gene expression. Am J Hum Genet **40:** 151-173.

HON, G. C., R. D. HAWKINS and B. REN, 2009 Predictive chromatin signatures in the mammalian genome. Hum Mol Genet **18:** R195-201.

HORE, T. A., J. E. DEAKIN and J. A. MARSHALL GRAVES, 2008 The evolution of epigenetic regulators CTCF and BORIS/CTCFL in amniotes. PLoS Genet **4:** e1000169.

HORE, T. A., R. W. RAPKINS and J. A. GRAVES, 2007 Construction and evolution of imprinted loci in mammals. Trends Genet **23:** 440-448.

HURST, L., 1997 *Genomic imprinting*. Oxford University Press, Oxford.

IAFRATE, A. J., L. FEUK, M. N. RIVERA, M. L. LISTEWNIK, P. K. DONAHOE *et al.*, 2004 Detection of large-scale variation in the human genome. Nat Genet **36:** 949-951.

IGLESIAS-PLATAS, I., D. MONK, J. JEBBINK, M. BUIMER, K. BOER *et al.*, 2007 STOX1 is not imprinted and is not likely to be involved in preeclampsia. Nature genetics **39:** 279-280; author reply 280-271.

INTERNATIONAL HAPMAP, C., 2005 A haplotype map of the human genome. Nature **437:** 1299-1320.

INTERNATIONAL HAPMAP, C., K. A. FRAZER, D. G. BALLINGER, D. R. COX, D. A. HINDS *et al.*, 2007 A second generation human haplotype map of over 3.1 million SNPs. Nature **449:** 851-861.

ISHIDA, M., D. MONK, A. J. DUNCAN, S. ABU-AMERO, J. CHONG *et al.*, 2012 Maternal Inheritance of a Promoter Variant in the Imprinted PHLDA2 Gene Significantly Increases Birth Weight. Am J Hum Genet **90:** 715-719.

JACOBS, P. A., C. M. WILSON, J. A. SPRENKLE, N. B. ROSENSHEIN and B. R. MIGEON, 1980 Mechanism of origin of complete hydatidiform moles. Nature **286:** 714-716.

JELINIC, P., J. C. STEHLE and P. SHAW, 2006 The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation. PLoS Biol **4:** e355.

JERU, I., P. DUQUESNOY, T. FERNANDES-ALNEMRI, E. COCHET, J. W. YU *et al.*, 2008 Mutations in NALP12 cause hereditary periodic fever syndromes. Proc Natl Acad Sci U S A **105:** 1614-1619.

JIA, D., R. Z. JURKOWSKA, X. ZHANG, A. JELTSCH and X. CHENG, 2007 Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature **449:** 248-251.

JIN, B., Q. TAO, J. PENG, H. M. SOO, W. WU *et al.*, 2008 DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. Hum Mol Genet **17:** 690-709.

JIN, Y., C. M. MAILLOUX, K. GOWAN, S. L. RICCARDI, G. LABERGE *et al.*, 2007 NALP1 in vitiligo-associated multiple autoimmune disease. N Engl J Med **356:** 1216-1225.

JOUVENOT, Y., F. POIRIER, J. JAMI and A. PALDI, 1999 Biallelic transcription of Igf2 and H19 in individual cells suggests a post-transcriptional contribution to genomic imprinting. Curr Biol **9:** 1199-1202.

JUDSON, H., B. E. HAYWARD, E. SHERIDAN and D. T. BONTHRON, 2002 A global disorder of imprinting in the human female germ line. Nature **416:** 539-542.

KACEM, S., and R. FEIL, 2009 Chromatin mechanisms in genomic imprinting. Mamm Genome **20:** 544-556.

KAGITANI, F., Y. KUROIWA, S. WAKANA, T. SHIROISHI, N. MIYOSHI *et al.*, 1997 Peg5/Neuronatin is an imprinted gene located on sub-distal chromosome 2 in the mouse. Nucleic Acids Res **25:** 3428-3432.

KAJII, T., and K. OHAMA, 1977 Androgenetic origin of hydatidiform mole. Nature **268:** 633-634.

KAMIYA, M., H. JUDSON, Y. OKAZAKI, M. KUSAKABE, M. MURAMATSU *et al.*, 2000 The cell cycle control gene ZAC/PLAGL1 is imprinted--a strong candidate gene for transient neonatal diabetes. Hum Mol Genet **9:** 453-460.

KANEDA, M., M. OKANO, K. HATA, T. SADO, N. TSUJIMOTO *et al.*, 2004 Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature **429:** 900-903.

KANEKO-ISHINO, T., Y. KUROIWA, N. MIYOSHI, T. KOHDA, R. SUZUKI *et al.*, 1995 Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. Nat Genet **11:** 52-59.

KANEKO-ISHINO, T., KUROIWA, Y., KHODA, T., SURANI, M.A. AND ISHINO, F., 1997 Systematic approaches for the identification of imprinted genes, pp. 146-164 in *Frontiers In Molecular Biology: Genomic Imprinting*, edited by M. A. A. R. SURANI, W. IRL Press, Oxford, UK.

KE, X., N. S. THOMAS, D. O. ROBINSON and A. COLLINS, 2002 The distinguishing sequence characteristics of mouse imprinted genes. Mamm Genome **13:** 639-645.

KELSEY, G., D. BODLE, H. J. MILLER, C. V. BEECHEY, C. COOMBES *et al.*, 1999 Identification of imprinted loci by methylation-sensitive representational difference analysis: application to mouse distal chromosome 2. Genomics **62:** 129-138.

KHOSLA, S., W. DEAN, D. BROWN, W. REIK and R. FEIL, 2001 Culture of preimplantation mouse embryos affects fetal development and the expression of imprinted genes. Biol Reprod **64:** 918-926.

KILLIAN, J. K., J. C. BYRD, J. V. JIRTLE, B. L. MUNDAY, M. K. STOSKOPF *et al.*, 2000 M6P/IGF2R imprinting evolution in mammals. Mol Cell **5:** 707-716.

KILLIAN, J. K., C. M. NOLAN, N. STEWART, B. L. MUNDAY, N. A. ANDERSEN *et al.*, 2001 Monotreme IGF2 expression and ancestral origin of genomic imprinting. J Exp Zool **291:** 205-212.

KITSBERG, D., S. SELIG, M. BRANDEIS, I. SIMON, I. KESHET *et al.*, 1993 Allele-specific replication timing of imprinted gene regions. Nature **364:** 459-463.

KNIGHT, J. C., B. J. KEATING and D. P. KWIATKOWSKI, 2004 Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. Nature genetics **36:** 394-399.

KNIGHT, J. C., B. J. KEATING, K. A. ROCKETT and D. P. KWIATKOWSKI, 2003 In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nature genetics **33:** 469-475.

KNOLL, J. H., S. D. CHENG and M. LALANDE, 1994 Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. Nat Genet **6:** 41-46.

KOBAYASHI, H., T. SAKURAI, M. IMAI, N. TAKAHASHI, A. FUKUDA *et al.*, 2012 Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet **8:** e1002440.

KOBAYASHI, S., A. ISOTANI, N. MISE, M. YAMAMOTO, Y. FUJIHARA *et al.*, 2006 Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, Rhox5/Pem, at preimplantation stages. Curr Biol **16:** 166-172.

KOBAYASHI, S., T. KOHDA, N. MIYOSHI, Y. KUROIWA, K. AISAKA *et al.*, 1997 Human PEG1/MEST, an imprinted gene on chromosome 7. Hum Mol Genet **6:** 781-786.

KOBAYASHI, S., H. WAGATSUMA, R. ONO, H. ICHIKAWA, M. YAMAZAKI *et al.*, 2000 Mouse Peg9/Dlk1 and human PEG9/DLK1 are paternally expressed imprinted genes closely located to the maternally expressed imprinted genes: mouse Meg3/Gtl2 and human MEG3. Genes Cells **5:** 1029-1037.

KONG, A., V. STEINTHORSDOTTIR, G. MASSON, G. THORLEIFSSON, P. SULEM *et al.*, 2009 Parental origin of sequence variants associated with complex diseases. Nature **462:** 868-874.

KOU, Y. C., L. SHAO, H. H. PENG, R. ROSETTA, D. DEL GAUDIO *et al.*, 2008 A recurrent intragenic genomic duplication, other novel mutations in NLRP7 and imprinting defects in recurrent biparental hydatidiform moles. Mol Hum Reprod **14:** 33-40.

KUHN, K., S. C. BAKER, E. CHUDIN, M. H. LIEU, S. OESER *et al.*, 2004 A novel, high-performance random array platform for quantitative gene expression profiling. Genome research **14:** 2347-2356.

KUROIWA, Y., T. KANEKO-ISHINO, F. KAGITANI, T. KOHDA, L. L. LI *et al.*, 1996 Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. Nat Genet **12:** 186-190.

LAMBERTINI, L., A. I. DIPLAS, M. J. LEE, R. SPERLING, J. CHEN *et al.*, 2008 A sensitive functional assay reveals frequent loss of genomic imprinting in human placenta. Epigenetics **3:** 261-269.

LANDER, E. S., L. M. LINTON, B. BIRREN, C. NUSBAUM, M. C. ZODY *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860-921.

LAWSON, H. A., J. E. CADY, C. PARTRIDGE, J. B. WOLF, C. F. SEMENKOVICH *et al.*, 2011 Genetic effects at pleiotropic Loci are context-dependent with consequences for the maintenance of genetic variation in populations. PLoS Genet **7:** e1002256.

LEE, C., A. J. IAFRATE and A. R. BROTHMAN, 2007 Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat Genet **39:** S48-54.

LEE, J., K. INOUE, R. ONO, N. OGONUKI, T. KOHDA *et al.*, 2002 Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development **129:** 1807-1817.

LEE, J. T., L. S. DAVIDOW and D. WARSHAWSKY, 1999a Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet **21:** 400-404.

LEE, M. P., S. BRANDENBURG, G. M. LANDES, M. ADAMS, G. MILLER *et al.*, 1999b Two novel genes in the center of the 11p15 imprinted domain escape genomic imprinting. Hum Mol Genet **8:** 683-690.

LEE, M. P., and A. P. FEINBERG, 1998 Genomic imprinting of a human apoptosis gene homologue, TSSC3. Cancer Res **58:** 1052-1056.

LEWIS, A., K. MITSUYA, D. UMLAUF, P. SMITH, W. DEAN *et al.*, 2004 Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet **36:** 1291-1295.

LEWIS, E. B., 1978 A gene complex controlling segmentation in Drosophila. Nature **276:** 565-570.

LI, E., C. BEARD and R. JAENISCH, 1993 Role for DNA methylation in genomic imprinting. Nature **366:** 362-365.

LI, E., T. H. BESTOR and R. JAENISCH, 1992 Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:** 915-926.

LI, J., A. J. BENCH, G. S. VASSILIOU, N. FOUROUCLAS, A. C. FERGUSON-SMITH *et al.*, 2004 Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies. Proc Natl Acad Sci U S A **101:** 7341-7346.

LI, J., R. SANTORO, K. KOBERNA and I. GRUMMT, 2005 The chromatin remodeling complex NoRC controls replication timing of rRNA genes. EMBO J **24:** 120-127.

LI, L. C., and R. DAHIYA, 2002 MethPrimer: designing primers for methylation PCRs. Bioinformatics **18:** 1427-1431.

LI, X., M. ITO, F. ZHOU, N. YOUNGSON, X. ZUO *et al.*, 2008 A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell **15:** 547-557.

LIDEGAARD, O., A. PINBORG and A. N. ANDERSEN, 2005 Imprinting diseases and IVF: Danish National IVF cohort study. Hum Reprod **20:** 950-954.

LIN, W., H. H. YANG and M. P. LEE, 2005 Allelic variation in gene expression identified through computational analysis of the dbEST database. Genomics **86:** 518-527.

LIPPMAN, Z., and R. MARTIENSSEN, 2004 The role of RNA interference in heterochromatic silencing. Nature **431:** 364-370.

LISTER, R., M. PELIZZOLA, R. H. DOWEN, R. D. HAWKINS, G. HON *et al.*, 2009 Human DNA methylomes at base resolution show widespread epigenomic differences. Nature.

LO, H. S., Z. WANG, Y. HU, H. H. YANG, S. GERE *et al.*, 2003 Allelic variation in gene expression is common in the human genome. Genome Research **13:** 1855-1862.

LO, Y. M., N. B. TSUI, R. W. CHIU, T. K. LAU, T. N. LEUNG *et al.*, 2007 Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat Med **13:** 218-223.

LUCIFERO, D., C. MERTINEIT, H. J. CLARKE, T. H. BESTOR and J. M. TRASLER, 2002 Methylation dynamics of imprinted genes in mouse germ cells. Genomics **79:** 530-538.

LUEDI, P. P., F. S. DIETRICH, J. R. WEIDMAN, J. M. BOSKO, R. L. JIRTLE *et al.*, 2007 Computational and experimental identification of novel human imprinted genes. Genome Res **17:** 1723-1730.

LUEDI, P. P., A. J. HARTEMINK and R. L. JIRTLE, 2005 Genome-wide prediction of imprinted murine genes. Genome research **15:** 875-884.

LUPSKI, J. R., 2007 Genomic rearrangements and sporadic disease. Nat Genet **39:** S43-47.

LYON, M. F., 1961 Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature **190:** 372-373.

MA, D., J. P. SHIELD, W. DEAN, I. LECLERC, C. KNAUF *et al.*, 2004 Impaired glucose homeostasis in transgenic mice expressing the human transient neonatal diabetes mellitus locus, TNDM. J Clin Invest **114:** 339-348.

MACKAY, D. J., J. L. CALLAWAY, S. M. MARKS, H. E. WHITE, C. L. ACERINI *et al.*, 2008 Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet **40:** 949-951.

MAHER, E. R., 2005 Imprinting and assisted reproductive technology. Human molecular genetics **14 Spec No 1:** R133-138.

MAHER, E. R., L. A. BRUETON, S. C. BOWDIN, A. LUHARIA, W. COOPER *et al.*, 2003 Beckwith-Wiedemann syndrome and assisted reproduction technology (ART). J Med Genet **40:** 62-64.

MANN, J. R., and R. H. LOVELL-BADGE, 1984 Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm. Nature **310:** 66-67.

MANN, M. R., S. S. LEE, A. S. DOHERTY, R. I. VERONA, L. D. NOLEN *et al.*, 2004 Selective loss of imprinting in the placenta following preimplantation development in culture. Development **131:** 3727-3735.

MARAGANORE, D. M., M. DE ANDRADE, T. G. LESNICK, K. J. STRAIN, M. J. FARRER *et al.*, 2005 High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet **77:** 685-693.

MARGUERON, R., P. TROJER and D. REINBERG, 2005 The key to development: interpreting the histone code? Curr Opin Genet Dev **15:** 163-176.

MARKS, H., J. C. CHOW, S. DENISSOV, K. J. FRANCOIJS, N. BROCKDORFF *et al.*, 2009 High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res **19:** 1361-1373.

MARQUES, C. J., F. CARVALHO, M. SOUSA and A. BARROS, 2004 Genomic imprinting in disruptive spermatogenesis. Lancet **363:** 1700-1702.

MARTIN, C., and Y. ZHANG, 2005 The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol **6:** 838-849.

MARTIN, C., and Y. ZHANG, 2007 Mechanisms of epigenetic inheritance. Curr Opin Cell Biol **19:** 266-272.

MAYER, C., K. M. SCHMITZ, J. LI, I. GRUMMT and R. SANTORO, 2006 Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol Cell **22:** 351-361.

MAYNARD, N. D., J. CHEN, R. K. STUART, J. B. FAN and B. REN, 2008 Genome-wide mapping of allele-specific protein-DNA interactions in human cells. Nat Methods **5:** 307-309.

MCCARROLL, S. A., and D. M. ALTSHULER, 2007 Copy-number variation and association studies of human disease. Nature genetics **39:** S37-42.

MCGRATH, J., and D. SOLTER, 1983 Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. Science **220:** 1300-1302.

MCGRATH, J., and D. SOLTER, 1984 Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro. Science **226:** 1317-1319.

MCGRATH, J., and D. SOLTER, 1986 Nuclear and cytoplasmic transfer in mammalian embryos. Developmental biology (New York, N.Y.: 1985) **4:** 37-55.

MCKENNA, E. S., and C. W. ROBERTS, 2009 Epigenetics and cancer without genomic instability. Cell Cycle **8:** 23-26.

MCMINN, J., M. WEI, N. SCHUPF, J. CUSMAI, E. B. JOHNSON *et al.*, 2006 Unbalanced Placental Expression of Imprinted Genes in Human Intrauterine Growth Restriction. Placenta **27:** 540-549.

MEFFORD, H. C., G. M. COOPER, T. ZERR, J. D. SMITH, C. BAKER *et al.*, 2009 A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. Genome Res **19:** 1579-1585.

MEFFORD, H. C., A. J. SHARP, C. BAKER, A. ITSARA, Z. JIANG *et al.*, 2008 Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. N Engl J Med **359:** 1685-1699.

MEGURO, M., A. KASHIWAGI, K. MITSUYA, M. NAKAO, I. KONDO *et al.*, 2001 A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome. Nat Genet **28:** 19-20.

MEIBOOM, M., H. MURUA ESCOBAR, F. PENTIMALLI, A. FUSCO, G. BELGE *et al.*, 2003 A 3.4-kbp transcript of ZNF331 is solely expressed in follicular thyroid adenomas. Cytogenet Genome Res **101:** 113-117.

MEYER, E., D. LIM, S. PASHA, L. J. TEE, F. RAHMAN *et al.*, 2009 Germline mutation in NLRP2 (NALP2) in a familial imprinting disorder (Beckwith-Wiedemann Syndrome). PLoS Genet **5:** e1000423.

MILANI, L., M. GUPTA, M. ANDERSEN, S. DHAR, M. FRYKNAS *et al.*, 2007 Allelic imbalance in gene expression as a guide to cis-acting regulatory single nucleotide polymorphisms in cancer cells. Nucleic Acids Res **35:** e34.

MIYOSHI, N., Y. KUROIWA, T. KOHDA, H. SHITARA, H. YONEKAWA *et al.*, 1998 Identification of the Meg1/Grb10 imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. Proc Natl Acad Sci U S A **95:** 1102-1107.

MIYOSHI, N., H. WAGATSUMA, S. WAKANA, T. SHIROISHI, M. NOMURA *et al.*, 2000 Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. Genes Cells **5:** 211-220.

MIZUNO, Y., Y. SOTOMARU, Y. KATSUZAWA, T. KONO, M. MEGURO *et al.*, 2002 Asb4, Ata3, and Dcn are novel imprinted genes identified by high-

throughput screening using RIKEN cDNA microarray. Biochem Biophys Res Commun **290:** 1499-1505.

MONK, D., P. ARNAUD, S. APOSTOLIDOU, F. A. HILLS, G. KELSEY *et al.*, 2006 Limited evolutionary conservation of imprinting in the human placenta. Proc Natl Acad Sci U S A **103:** 6623-6628.

MONK, D., A. WAGSCHAL, P. ARNAUD, P. MULLER, L. PARKER-KATIRAEE *et al.*, 2008 Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires G9a and Eed for allelic-silencing. Genome Res.

MONTGOMERY, S. B., M. SAMMETH, M. GUTIERREZ-ARCELUS, R. P. LACH, C. INGLE *et al.*, 2010 Transcriptome genetics using second generation sequencing in a Caucasian population. Nature **464:** 773-777.

MOORE, T., and D. HAIG, 1991 Genomic imprinting in mammalian development: a parental tug-of-war. Trends in genetics : TIG **7:** 45-49.

MORCOS, L., B. GE, V. KOKA, K. C. LAM, D. K. POKHOLOK *et al.*, 2011 Genome-wide assessment of imprinted expression in human cells. Genome Biol **12:** R25.

MORGAN, H. D., X. L. JIN, A. LI, E. WHITELAW and C. O'NEILL, 2008 The culture of zygotes to the blastocyst stage changes the postnatal expression of an epigentically labile allele, agouti viable yellow, in mice. Biol Reprod **79:** 618-623.

MORISON, I. M., J. P. RAMSAY and H. G. SPENCER, 2005 A census of mammalian imprinting. Trends in genetics : TIG **21:** 457-465.

MORISON, I. M., and A. E. REEVE, 1998 A catalogue of imprinted genes and parent-of-origin effects in humans and animals. Human molecular genetics **7:** 1599-1609.

MORLEY, M., C. M. MOLONY, T. M. WEBER, J. L. DEVLIN, K. G. EWENS *et al.*, 2004 Genetic analysis of genome-wide variation in human gene expression. Nature **430:** 743-747.

MOSTOSLAVSKY, R., N. SINGH, T. TENZEN, M. GOLDMIT, C. GABAY *et al.*, 2001 Asynchronous replication and allelic exclusion in the immune system. Nature **414:** 221-225.

MULLIS, K., F. FALOONA, S. SCHARF, R. SAIKI, G. HORN *et al.*, 1986 Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol **51 Pt 1:** 263-273.

MULLIS, K. B., and F. A. FALOONA, 1987 Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol **155:** 335-350.

MURDOCH, S., U. DJURIC, B. MAZHAR, M. SEOUD, R. KHAN *et al.*, 2006 Mutations in NALP7 cause recurrent hydatidiform moles and reproductive wastage in humans. Nat Genet **38:** 300-302.

MURPHY, S. K., and R. L. JIRTLE, 2003 Imprinting evolution and the price of silence. Bioessays **25:** 577-588.

MURPHY, S. K., A. A. WYLIE and R. L. JIRTLE, 2001 Imprinting of PEG3, the human homologue of a mouse gene involved in nurturing behavior. Genomics **71:** 110-117.

NAVARRO, P., S. PICHARD, C. CIAUDO, P. AVNER and C. ROUGEULLE, 2005 Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. Genes Dev **19:** 1474-1484.

NICA, A. C., S. B. MONTGOMERY, A. S. DIMAS, B. E. STRANGER, C. BEAZLEY *et al.*, 2010 Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet **6:** e1000895.

NICA, A. C., L. PARTS, D. GLASS, J. NISBET, A. BARRETT *et al.*, 2011 The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet **7:** e1002003.

NICHOLLS, R. D., J. H. KNOLL, M. G. BUTLER, S. KARAM and M. LALANDE, 1989 Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. Nature **342:** 281-285.

NIKAIDO, I., C. SAITO, Y. MIZUNO, M. MEGURO, H. BONO *et al.*, 2003 Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. Genome research **13:** 1402-1409.

NOGUER-DANCE, M., S. ABU-AMERO, M. AL-KHTIB, A. LEFEVRE, P. COULLIN *et al.*, 2010 The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. Hum Mol Genet **19:** 3566-3582.

NOLAN, C. M., J. K. KILLIAN, J. N. PETITTE and R. L. JIRTLE, 2001 Imprint status of M6P/IGF2R and IGF2 in chickens. Dev Genes Evol **211:** 179-183.

O'NEILL, M. J., R. S. INGRAM, P. B. VRANA and S. M. TILGHMAN, 2000 Allelic expression of IGF2 in marsupials and birds. Dev Genes Evol **210:** 18-20.

OBATA, Y., T. KANEKO-ISHINO, T. KOIDE, Y. TAKAI, T. UEDA *et al.*, 1998 Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesis. Development **125:** 1553-1560.

OETH, P., G. DEL MISTRO, G. MARNELLOS, T. SHI and D. VAN DEN BOOM, 2009 Qualitative and quantitative genotyping using single base primer

extension coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MassARRAY). Methods Mol Biol **578:** 307-343.

OHAMA, K., T. KAJII, E. OKAMOTO, Y. FUKUDA, K. IMAIZUMI *et al.*, 1981 Dispermic origin of XY hydatidiform moles. Nature **292:** 551-552.

OHLSSON, R., A. NYSTROM, S. PFEIFER-OHLSSON, V. TOHONEN, F. HEDBORG *et al.*, 1993 IGF2 is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. Nature genetics **4:** 94-97.

OHLSSON, R., B. TYCKO and C. SAPIENZA, 1998 Monoallelic expression: 'there can only be one'. Trends Genet **14:** 435-438.

OHNO, S., and T. S. HAUSCHKA, 1960 Allocycly of the X-chromosome in tumors and normal tissues. Cancer Res **20:** 541-545.

OKAMURA, K., Y. HAGIWARA-TAKEUCHI, T. LI, T. H. VU, M. HIRAI *et al.*, 2000 Comparative genome analysis of the mouse imprinted gene impact and its nonimprinted human homolog IMPACT: toward the structural basis for species-specific imprinting. Genome Res **10:** 1878-1889.

OKANO, M., D. W. BELL, D. A. HABER and E. LI, 1999 DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99:** 247-257.

OKUTSU, T., Y. KUROIWA, F. KAGITANI, M. KAI, K. AISAKA *et al.*, 2000 Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. J Biochem **127:** 475-483.

ONO, R., S. KOBAYASHI, H. WAGATSUMA, K. AISAKA, T. KOHDA *et al.*, 2001 A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. Genomics **73:** 232-237.

ORSTAVIK, K. H., 2009 X chromosome inactivation in clinical practice. Hum Genet **126:** 363-373.

ORSTAVIK, K. H., K. EIKLID, C. B. VAN DER HAGEN, S. SPETALEN, K. KIERULF *et al.*, 2003 Another case of imprinting defect in a girl with Angelman syndrome who was conceived by intracytoplasmic semen injection. Am J Hum Genet **72:** 218-219.

OSBORNE, C. S., L. CHAKALOVA, K. E. BROWN, D. CARTER, A. HORTON *et al.*, 2004 Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet **36:** 1065-1071.

PALOMAKI, G. E., E. M. KLOZA, G. M. LAMBERT-MESSERLIAN, J. E. HADDOW, L. M. NEVEUX *et al.*, 2011 DNA sequencing of maternal plasma to detect

Down syndrome: an international clinical validation study. Genet Med **13:** 913-920.

PANT, P. V., H. TAO, E. J. BEILHARZ, D. G. BALLINGER, D. R. COX *et al.*, 2006 Analysis of allelic differential expression in human white blood cells. Genome research.

PARRY, D. A., C. V. LOGAN, B. E. HAYWARD, M. SHIRES, H. LANDOLSI *et al.*, 2011 Mutations causing familial biparental hydatidiform mole implicate c6orf221 as a possible regulator of genomic imprinting in the human oocyte. Am J Hum Genet **89:** 451-458.

PASTINEN, T., and T. J. HUDSON, 2004 Cis-acting regulatory variation in the human genome. Science **306:** 647-650.

PASTINEN, T., R. SLADEK, S. GURD, A. SAMMAK, B. GE *et al.*, 2004 A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics **16:** 184-193.

PERNIS, B., G. CHIAPPINO, A. S. KELUS and P. G. GELL, 1965 Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. J Exp Med **122:** 853-876.

PETERS, J., S. F. WROE, C. A. WELLS, H. J. MILLER, D. BODLE *et al.*, 1999 A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. Proc Natl Acad Sci U S A **96:** 3830-3835.

PICKRELL, J. K., J. C. MARIONI, A. A. PAI, J. F. DEGNER, B. E. ENGELHARDT *et al.*, 2010 Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature **464:** 768-772.

PLASS, C., H. SHIBATA, I. KALCHEVA, L. MULLINS, N. KOTELEVTSEVA *et al.*, 1996 Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat Genet **14:** 106-109.

POLLARD, K. S., D. SERRE, X. WANG, H. TAO, E. GRUNDBERG *et al.*, 2008 A genome-wide approach to identifying novel-imprinted genes. Hum Genet **122:** 625-634.

QIAN, N., D. FRANK, D. O'KEEFE, D. DAO, L. ZHAO *et al.*, 1997 The IPL gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to TDAG51, implicated in Fas expression and apoptosis. Hum Mol Genet **6:** 2021-2029.

RAEFSKI, A. S., and M. J. O'NEILL, 2005 Identification of a cluster of X-linked imprinted genes in mice. Nat Genet **37:** 620-624.

RAINIER, S., L. A. JOHNSON, C. J. DOBRY, A. J. PING, P. E. GRUNDY *et al.*, 1993 Relaxation of imprinted genes in human cancer. Nature **362:** 747-749.

RAJEWSKY, K., 1996 Clonal selection and learning in the antibody system. Nature **381:** 751-758.

RAZIN, A., and H. CEDAR, 1994 DNA methylation and genomic imprinting. Cell **77:** 473-476.

REDON, R., S. ISHIKAWA, K. R. FITCH, L. FEUK, G. H. PERRY *et al.*, 2006 Global variation in copy number in the human genome. Nature **444:** 444-454.

REIK, W., 2007 Stability and flexibility of epigenetic gene regulation in mammalian development. Nature **447:** 425-432.

REIK, W., and J. WALTER, 2001 Genomic imprinting: parental influence on the genome. Nature reviews.Genetics **2:** 21-32.

RENFREE, M. B., T. A. HORE, G. SHAW, J. A. GRAVES and A. J. PASK, 2009 Evolution of genomic imprinting: insights from marsupials and monotremes. Annu Rev Genomics Hum Genet **10:** 241-262.

RHOADES, K. L., N. SINGH, I. SIMON, B. GLIDDEN, H. CEDAR *et al.*, 2000 Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. Curr Biol **10:** 789-792.

RIPPE, V., G. BELGE, M. MEIBOOM, B. KAZMIERCZAK, A. FUSCO *et al.*, 1999 A KRAB zinc finger protein gene is the potential target of 19q13 translocation in benign thyroid tumors. Genes Chromosomes Cancer **26:** 229-236.

RITCHIE, M. E., M. S. FORREST, A. S. DIMAS, C. DAELEMANS, E. T. DERMITZAKIS *et al.*, 2010 Data analysis issues for allele-specific expression using Illumina's GoldenGate assay. BMC Bioinformatics **11:** 280.

RIVERA, R. M., P. STEIN, J. R. WEAVER, J. MAGER, R. M. SCHULTZ *et al.*, 2008 Manipulations of mouse embryos prior to implantation result in aberrant expression of imprinted genes on day 9.5 of development. Hum Mol Genet **17:** 1-14.

RODRIGUEZ, I., P. FEINSTEIN and P. MOMBAERTS, 1999 Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. Cell **97:** 199-208.

ROHDE, C., Y. ZHANG, T. P. JURKOWSKI, H. STAMERJOHANNS, R. REINHARDT *et al.*, 2008 Bisulfite sequencing Data Presentation and Compilation (BDPC) web server--a useful tool for DNA methylation analysis. Nucleic Acids Res **36:** e34.

ROSSANT, J., and J. C. CROSS, 2001 Placental development: lessons from mouse mutants. Nat Rev Genet **2:** 538-548.

ROY, S., J. ERNST, P. V. KHARCHENKO, P. KHERADPOUR, N. NEGRE *et al.*, 2010 Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science **330:** 1787-1797.

ROZEN, S., and H. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132:** 365-386.

RUF, N., S. BAHRING, D. GALETZKA, G. PLIUSHCH, F. C. LUFT *et al.*, 2007 Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. Human molecular genetics **16:** 2591-2599.

RUF, N., U. DUNZINGER, A. BRINCKMANN, T. HAAF, P. NURNBERG *et al.*, 2006 Expression profiling of uniparental mouse embryos is inefficient in identifying novel imprinted genes. Genomics **87:** 509-519.

SACHIDANANDAM, R., D. WEISSMAN, S. C. SCHMIDT, J. M. KAKOL, L. D. STEIN *et al.*, 2001 A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409:** 928-933.

SADO, T., Y. HOKI and H. SASAKI, 2005 Tsix silences Xist through modification of chromatin structure. Dev Cell **9:** 159-165.

SADRI, R., and P. J. HORNSBY, 1996 Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. Nucleic Acids Res **24:** 5058-5059.

SAGARA, J., T. HIGUCHI, Y. HATTORI, M. MORIYA, H. SARVOTHAM *et al.*, 2003 Scapinin, a putative protein phosphatase-1 regulatory subunit associated with the nuclear nonchromatin structure. J Biol Chem **278:** 45611-45619.

SAIKI, R. K., T. L. BUGAWAN, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1986 Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature **324:** 163-166.

SAKATANI, T., M. WEI, M. KATOH, C. OKITA, D. WADA *et al.*, 2001 Epigenetic heterogeneity at imprinted loci in normal populations. Biochemical and biophysical research communications **283:** 1124-1130.

SANTORO, R., and I. GRUMMT, 2001 Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. Mol Cell **8:** 719-725.

SANTORO, R., K. M. SCHMITZ, J. SANDOVAL and I. GRUMMT, 2010 Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. EMBO Rep **11:** 52-58.

SANTOS, F., L. HYSLOP, P. STOJKOVIC, C. LEARY, A. MURDOCH *et al.*, 2010 Evaluation of epigenetic marks in human embryos derived from IVF and ICSI. Hum Reprod **25:** 2387-2395.

SAPIENZA, C., A. C. PETERSON, J. ROSSANT and R. BALLING, 1987 Degree of methylation of transgenes is dependent on gamete of origin. Nature **328:** 251-254.

SASAKI, H., T. HAMADA, T. UEDA, R. SEKI, T. HIGASHINAKAGAWA *et al.*, 1991 Inherited type of allelic methylation variations in a mouse chromosome region where an integrated transgene shows methylation imprinting. Development **111:** 573-581.

SCHAEFER, C. B., S. K. OOI, T. H. BESTOR and D. BOURC'HIS, 2007 Epigenetic decisions in mammalian germ cells. Science **316:** 398-399.

SCHALKWYK, L. C., E. L. MEABURN, R. SMITH, E. L. DEMPSTER, A. R. JEFFRIES *et al.*, 2010 Allelic skewing of DNA methylation is widespread across the genome. Am J Hum Genet **86:** 196-212.

SCHERER, S. W., C. LEE, E. BIRNEY, D. M. ALTSHULER, E. E. EICHLER *et al.*, 2007 Challenges and standards in integrating surveys of structural variation. Nat Genet **39:** S7-15.

SCHIEVE, L. A., S. F. MEIKLE, C. FERRE, H. B. PETERSON, G. JENG *et al.*, 2002 Low and very low birth weight in infants conceived with use of assisted reproductive technology. The New England journal of medicine **346:** 731-737.

SCHLESINGER, S., S. SELIG, Y. BERGMAN and H. CEDAR, 2009 Allelic inactivation of rDNA loci. Genes Dev **23:** 2437-2447.

SCHMIDT, J. V., P. G. MATTESON, B. K. JONES, X. J. GUAN and S. M. TILGHMAN, 2000 The Dlk1 and Gtl2 genes are linked and reciprocally imprinted. Genes Dev **14:** 1997-2002.

SCHOENHERR, C. J., J. M. LEVORSE and S. M. TILGHMAN, 2003 CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet **33:** 66-69.

SCHULZ, R., T. R. MENHENIOTT, K. WOODFINE, A. J. WOOD, J. D. CHOI *et al.*, 2006 Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. Nucleic Acids Res **34:** e88.

SEBAT, J., B. LAKSHMI, D. MALHOTRA, J. TROGE, C. LESE-MARTIN *et al.*, 2007 Strong association of de novo copy number mutations with autism. Science **316:** 445-449.

SEISENBERGER, S., S. ANDREWS, F. KRUEGER, J. ARAND, J. WALTER *et al.*, 2012 The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell **48:** 849-862.

SEISENBERGER, S., J. R. PEAT, T. A. HORE, F. SANTOS, W. DEAN *et al.*, 2013a Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. Philos Trans R Soc Lond B Biol Sci **368:** 20110330.

SEISENBERGER, S., J. R. PEAT and W. REIK, 2013b Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. Curr Opin Cell Biol.

SEOIGHE, C., V. NEMBAWARE and K. SCHEFFLER, 2006 Maximum likelihood inference of imprinting and allele-specific expression from EST data. Bioinformatics (Oxford, England) **22:** 3032-3039.

SERRE, D., S. GURD, B. GE, R. SLADEK, D. SINNETT *et al.*, 2008a Differential Allelic Expression in the Human Genome: A Robust Approach to Identify Genetic and Epigenetic Cis-Acting Mechanisms Regulating Gene Expression. PLoS Genetics.

SERRE, D., S. GURD, B. GE, R. SLADEK, D. SINNETT *et al.*, 2008b Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. PLoS Genet **4:** e1000006.

SHARP, A. J., E. STATHAKI, E. MIGLIAVACCA, M. BRAHMACHARY, S. B. MONTGOMERY *et al.*, 2011 DNA methylation profiles of human active and inactive X chromosomes. Genome Res **21:** 1592-1600.

SMALL, K. S., A. K. HEDMAN, E. GRUNDBERG, A. C. NICA, G. THORLEIFSSON *et al.*, 2011 Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nat Genet **43:** 561-564.

SMALLWOOD, S. A., S. TOMIZAWA, F. KRUEGER, N. RUF, N. CARLI *et al.*, 2011 Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet **43:** 811-814.

SMITH, G. C., E. J. STENHOUSE, J. A. CROSSLEY, D. A. AITKEN, A. D. CAMERON *et al.*, 2002 Early-pregnancy origins of low birth weight. Nature **417:** 916.

SMITH, R. J., P. ARNAUD and G. KELSEY, 2004 Identification and properties of imprinted genes and their control elements. Cytogenet Genome Res **105:** 335-345.

SMITH, R. J., W. DEAN, G. KONFORTOVA and G. KELSEY, 2003 Identification of novel imprinted genes in a genome-wide screen for maternal methylation. Genome Res **13:** 558-569.

SMITH, Z. D., M. M. CHAN, T. S. MIKKELSEN, H. GU, A. GNIRKE *et al.*, 2012 A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature **484:** 339-344.

SMITS, G., A. J. MUNGALL, S. GRIFFITHS-JONES, P. SMITH, D. BEURY *et al.*, 2008 Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. Nat Genet **40:** 971-976.

SMRZKA, O. W., I. FAE, R. STOGER, R. KURZBAUER, G. F. FISCHER *et al.*, 1995 Conservation of a maternal-specific methylation signal at the human IGF2R locus. Hum Mol Genet **4:** 1945-1952.

SMYTH, G. K., 2004 Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** Article3.

SOOD, R., J. L. ZEHNDER, M. L. DRUZIN and P. O. BROWN, 2006 Gene expression patterns in human placenta. Proc Natl Acad Sci U S A **103:** 5478-5483.

SPAHN, L., and D. P. BARLOW, 2003 An ICE pattern crystallizes. Nat Genet **35:** 11-12.

SPIELMAN, R. S., L. A. BASTONE, J. T. BURDICK, M. MORLEY, W. J. EWENS *et al.*, 2007 Common genetic variants account for differences in gene expression among ethnic groups. Nat Genet **39:** 226-231.

STANSSENS, P., M. ZABEAU, G. MEERSSEMAN, G. REMES, Y. GANSEMANS *et al.*, 2004 High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. Genome research **14:** 126-133.

STEFANSSON, H., D. RUJESCU, S. CICHON, O. P. PIETILAINEN, A. INGASON *et al.*, 2008 Large recurrent microdeletions associated with schizophrenia. Nature **455:** 232-236.

STRAIN, L., J. P. WARNER, T. JOHNSTON and D. T. BONTHRON, 1995 A human parthenogenetic chimaera. Nat Genet **11:** 164-169.

STRANGER, B. E., M. S. FORREST, A. G. CLARK, M. J. MINICHIELLO, S. DEUTSCH *et al.*, 2005 Genome-wide associations of gene expression variation in humans. PLoS genetics **1:** e78.

STRANGER, B. E., M. S. FORREST, M. DUNNING, C. E. INGLE, C. BEAZLEY *et al.*, 2007 Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science (New York, N.Y.) **315:** 848-853.

STRICHMAN-ALMASHANU, L. Z., R. S. LEE, P. O. ONYANGO, E. PERLMAN, F. FLAM *et al.*, 2002 A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. Genome research **12:** 543-554.

SUN, B. K., A. M. DEATON and J. T. LEE, 2006 A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell **21:** 617-628.

SURANI, M. A., S. C. BARTON and M. L. NORRIS, 1984 Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308:** 548-550.

SUZUKI, S., M. B. RENFREE, A. J. PASK, G. SHAW, S. KOBAYASHI *et al.*, 2005 Genomic imprinting of IGF2, p57(KIP2) and PEG1/MEST in a marsupial, the tammar wallaby. Mech Dev **122:** 213-222.

SWAIN, J. L., T. A. STEWART and P. LEDER, 1987 Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50:** 719-727.

SZABO, P., S. H. TANG, A. RENTSENDORJ, G. P. PFEIFER and J. R. MANN, 2000 Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. Curr Biol **10:** 607-610.

SZABO, P. E., and J. R. MANN, 1995 Biallelic expression of imprinted genes in the mouse germ line: implications for erasure, establishment, and mechanisms of genomic imprinting. Genes Dev **9:** 1857-1868.

TAKADA, S., M. TEVENDALE, J. BAKER, P. GEORGIADES, E. CAMPBELL *et al.*, 2000 Delta-like and gtl2 are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. Curr Biol **10:** 1135-1138.

TAKAGI, N., and M. SASAKI, 1975 Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature **256:** 640-642.

TAN, A. C., J. B. FAN, C. KARIKARI, M. BIBIKOVA, E. W. GARCIA *et al.*, 2007 Allele-specific expression in the germline of patients with familial pancreatic cancer: An unbiased approach to cancer gene discovery. Cancer Biol Ther **7**.

TAN, A. C., J. B. FAN, C. KARIKARI, M. BIBIKOVA, E. W. GARCIA *et al.*, 2008 Allele-specific expression in the germline of patients with familial pancreatic cancer: an unbiased approach to cancer gene discovery. Cancer Biol Ther **7:** 135-144.

TAYLOR, J. H., 1960 Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. J Biophys Biochem Cytol **7:** 455-464.

THE INTERNATIONAL HAPMAP, C., 2003 The International HapMap Project. Nature **426:** 789-796.

THIRIET, C., and J. J. HAYES, 2005 Chromatin in need of a fix: phosphorylation of H2AX connects chromatin to DNA repair. Mol Cell **18:** 617-622.

THORVALDSEN, J. L., and M. S. BARTOLOMEI, 2007 SnapShot: Imprinted Gene Clusters. Cell **130:** 958.

TSAI, K. W., H. W. KAO, H. C. CHEN, S. J. CHEN and W. C. LIN, 2009 Epigenetic control of the expression of a primate-specific microRNA cluster in human cancer cells. Epigenetics **4:** 587-592.

TYCKO, B., 2006 Imprinted genes in placental growth and obstetric disorders. Cytogenet Genome Res **113:** 271-278.

UBEDA, F., 2008 Evolution of genomic imprinting with biparental care: implications for Prader-Willi and Angelman syndromes. PLoS Biol **6:** e208.

UMLAUF, D., Y. GOTO, R. CAO, F. CERQUEIRA, A. WAGSCHAL *et al.*, 2004 Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat Genet **36:** 1296-1300.

VAN DEN VEYVER, I. B., B. NORMAN, C. Q. TRAN, J. BOURJAC and R. SLIM, 2001 The human homologue (PEG3) of the mouse paternally expressed gene 3 (Peg3) is maternally imprinted but not mutated in women with familial recurrent hydatidiform molar pregnancies. J Soc Gynecol Investig **8:** 305-313.

VAN DIJK, M., S. DREWLO and C. B. OUDEJANS, 2010 Differential methylation of STOX1 in human placenta. Epigenetics **5:** 736-742.

VARMUZA, S., and M. MANN, 1994 Genomic imprinting--defusing the ovarian time bomb. Trends Genet **10:** 118-123.

VARRAULT, A., C. GUEYDAN, A. DELALBRE, A. BELLMANN, S. HOUSSAMI *et al.*, 2006 Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. Dev Cell **11:** 711-722.

VENTER, J. C., M. D. ADAMS, E. W. MYERS, P. W. LI, R. J. MURAL *et al.*, 2001 The sequence of the human genome. Science (New York, N.Y.) **291:** 1304-1351.

VERONA, R. I., M. R. MANN and M. S. BARTOLOMEI, 2003 Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol **19:** 237-259.

WALLACE, C., D. J. SMYTH, M. MAISURIA-ARMER, N. M. WALKER, J. A. TODD *et al.*, 2010 The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. Nat Genet **42:** 68-71.

WALTER, J., and M. PAULSEN, 2003 Imprinting and disease. Semin Cell Dev Biol **14:** 101-110.

WANG, J., W. WANG, R. LI, Y. LI, G. TIAN *et al.*, 2008a The diploid genome sequence of an Asian individual. Nature **456:** 60-65.

WANG, X., P. D. SOLOWAY and A. G. CLARK, 2011 A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics **189:** 109-122.

WANG, X., Q. SUN, S. D. MCGRATH, E. R. MARDIS, P. D. SOLOWAY *et al.*, 2008b Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS ONE **3:** e3839.

WATANABE, D., and D. P. BARLOW, 1996 Random and imprinted monoallelic expression. Genes Cells **1:** 795-802.

WEN, B., H. WU, H. BJORNSSON, R. D. GREEN, R. IRIZARRY *et al.*, 2008 Overlapping euchromatin/heterochromatin- associated marks are enriched in imprinted gene regions and predict allele-specific modification. Genome Res **18:** 1806-1813.

WHEELER, D. A., M. SRINIVASAN, M. EGHOLM, Y. SHEN, L. CHEN *et al.*, 2008 The complete genome of an individual by massively parallel DNA sequencing. Nature **452:** 872-876.

WIDER, C., S. J. LINCOLN, M. G. HECKMAN, N. N. DIEHL, J. T. STONE *et al.*, 2009 Phactr2 and Parkinson's disease. Neurosci Lett **453:** 9-11.

WILKINS, J. F., and D. HAIG, 2003 What good is genomic imprinting: the function of parent-specific gene expression. Nature reviews.Genetics **4:** 359-368.

WINN, V. D., R. HAIMOV-KOCHMAN, A. C. PAQUET, Y. J. YANG, M. S. MADHUSUDHAN *et al.*, 2007 Gene expression profiling of the human maternal-fetal interface reveals dramatic changes between midgestation and term. Endocrinology **148:** 1059-1079.

WOLF, J. B., J. M. CHEVERUD, C. ROSEMAN and R. HAGER, 2008 Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. PLoS Genet **4:** e1000091.

WOOD, A. J., and R. J. OAKEY, 2006 Genomic imprinting in mammals: emerging themes and established theories. PLoS Genet **2:** e147.

WOOD, A. J., R. G. ROBERTS, D. MONK, G. E. MOORE, R. SCHULZ *et al.*, 2007 A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. PLoS Genet **3:** e20.

WU, H., S. ZHANG, W. QIU, G. ZHANG, Q. XIA *et al.*, 2001 Isolation, characterization, and mapping of a novel human KRAB zinc finger protein encoding gene ZNF463. Biochim Biophys Acta **1518:** 190-193.

WYLIE, A. A., S. K. MURPHY, T. C. ORTON and R. L. JIRTLE, 2000 Novel imprinted DLK1/GTL2 domain on human chromosome 14 contains motifs that mimic those implicated in IGF2/H19 regulation. Genome Res **10:** 1711-1718.

XIONG, Z., and P. W. LAIRD, 1997 COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res **25:** 2532-2534.

XU, G. L., T. H. BESTOR, D. BOURC'HIS, C. L. HSIEH, N. TOMMERUP *et al.*, 1999 Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature **402:** 187-191.

YAN, H., W. YUAN, V. E. VELCULESCU, B. VOGELSTEIN and K. W. KINZLER, 2002 Allelic variation in human gene expression. Science **297:** 1143.

YANG, H. H., Y. HU, M. EDMONSON, K. BUETOW and M. P. LEE, 2003 Computation method to identify differential allelic gene expression and novel imprinted genes. Bioinformatics **19:** 952-955.

YANG, T. P., C. BEAZLEY, S. B. MONTGOMERY, A. S. DIMAS, M. GUTIERREZ-ARCELUS *et al.*, 2010 Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics **26:** 2474-2476.

YASUDA, K., K. MIYAKE, Y. HORIKAWA, K. HARA, H. OSAWA *et al.*, 2008 Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet **40:** 1092-1097.

YEN, P. H., P. PATEL, A. C. CHINAULT, T. MOHANDAS and L. J. SHAPIRO, 1984 Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. Proc Natl Acad Sci U S A **81:** 1759-1763.

YIN, D., D. XIE, S. DE VOS, G. LIU, C. W. MILLER *et al.*, 2004 Imprinting status of DLK1 gene in brain tumors and lymphomas. Int J Oncol **24:** 1011-1015.

YOKOMINE, T., K. HATA, M. TSUDZUKI and H. SASAKI, 2006 Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. Cytogenet Genome Res **113:** 75-80.

YOON, B. J., H. HERMAN, A. SIKORA, L. T. SMITH, C. PLASS *et al.*, 2002 Regulation of DNA methylation of Rasgrf1. Nat Genet **30:** 92-96.

YOUNG, L. E., K. FERNANDES, T. G. MCEVOY, S. C. BUTTERWITH, C. G. GUTIERREZ *et al.*, 2001 Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture. Nat Genet **27:** 153-154.

ZHANG, A., D. A. SKAAR, Y. LI, D. HUANG, T. M. PRICE *et al.*, 2011 Novel retrotransposed imprinted locus identified at human 6p25. Nucleic Acids Res **39:** 5388-5400.

ZHANG, Y., 2003 Transcriptional regulation by histone ubiquitination and deubiquitination. Genes Dev **17:** 2733-2740.

ZHANG, Y., C. ROHDE, S. TIERLING, H. STAMERJOHANNS, R. REINHARDT *et al.*, 2009 DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. Methods Mol Biol **507:** 177-187.

ZHANG, Y., and B. TYCKO, 1992 Monoallelic expression of the human H19 gene. Nat Genet **1:** 40-44.

ZHAO, J., T. K. OHSUMI, J. T. KUNG, Y. OGAWA, D. J. GRAU *et al.*, 2010 Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell **40:** 939-953.

ZHAO, Z., G. TAVOOSIDANA, M. SJOLINDER, A. GONDOR, P. MARIANO *et al.*, 2006 Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet **38:** 1341-1347.

ZWART, R., F. SLEUTELS, A. WUTZ, A. H. SCHINKEL and D. P. BARLOW, 2001 Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev **15:** 2361-2366.

Appendices

Appendix 1: Genes and SNPs tested

Table 1: List of genes and SNPs tested on Sequenom platform

GENE	SNPs	Status	Status if successful (1)
ACAS2	rs4911163	Mouse candidate	1
ACOXL	rs7558938	Mouse candidate	1
ADAR	rs1127326	Human candidate	failed
ADPGK	rs9460	Mouse candidate	1
AGPAT5	rs2911970	Mouse candidate	1
ALDH1B1	rs2073478	Mouse candidate	1
AMPD3	rs3741041	Biallelic control	1
ANTXR2	rs7747	Mouse candidate	1
ATG16L	rs1045100	Mouse candidate	1
ARHGAP28	rs4239328	Mouse candidate	1
ATP10A	rs2066707	Imprinted control	no hets
ATP9B	rs3591	Mouse candidate	1
BCL2L11	rs6753785	Mouse candidate	1
BMP6	rs17557	Mouse candidate	1
BSG	rs8259	Human candidate	failed
C10orf9	rs1043583	Mouse candidate	1
C14orf100	rs7560	Mouse candidate	1
C16orf57	rs11551263	Mouse candidate	1
C18orf4	rs7227616	Mouse candidate	not expressed
C19orf48	rs9991	Human candidate	1
C19orf6	rs7146	Mouse candidate	1
C1orf121	rs2242448	Mouse candidate	1
C1orf164	rs1051664	Mouse candidate	1
C22orf25	rs737986	Mouse candidate	1
C9orf93	rs1539172	Mouse candidate	1
CALCR	rs1801197	Imprinted in mouse	1
CARD9	rs10781499	Mouse candidate	1
CAST	rs754615	Mouse candidate	1
CCDC86	rs7167	Mouse candidate	1
CDCA1	rs1509022	Mouse candidate	1
CEP72	rs2458815	Mouse candidate	failed
CGI-69	rs2011951	Human candidate	1
ChGn	rs6984644	Mouse candidate	failed
COASY	rs615942	Human candidate	1
COL18A1	rs7499	Mouse candidate	1
COL1A2	rs1060399	Human candidate	failed

COMT	rs4633	Human candidate	1
COQ7	rs11074359	Mouse candidate	1
CPXM2	rs10794567	Mouse candidate	1
CRYAA	rs872331	Human candidate	not expressed
CRYZ	rs17459	Mouse candidate	1
CTSD	rs12214	Human and mouse candidate	1
CYRR1	rs2830239	Mouse candidate	1
DCHS1	rs997263	Mouse candidate	1
DDAH1	rs233112	Mouse candidate	1
DDIT4L	rs1053227	Mouse candidate	1
DHCR7	rs1790345	Biallelic control	1
DISC1	rs821616	Mouse candidate	1
DLK1	rs1802710	Imprinted control	1
DOCK5	rs2271108	Mouse candidate	1
EDNRA	rs5333	Mouse candidate	1
EMILIN3	rs6072352	Mouse candidate	1
ERCC5	rs1047768	Human candidate	1
FGB	rs6056	Human candidate	1
FMO4	rs1042772	Mouse candidate	1
FMOD	rs4605	Human candidate	1
FOSL2	rs7562	Mouse candidate	1
GAPDH	rs1803622	Biallelic control	1
GATM	rs1145086	Imprinted in mouse	1
GNAI3	rs2301230	Mouse candidate	1
GPR158	rs10828833	Mouse candidate	1
GRIA1	rs707176	Mouse candidate	not expressed
GUSB	rs9530	Mouse candidate	1
HCA112	rs9088	Mouse candidate	1
HCLS1	rs1128163	Mouse candidate	1
HES6	rs9776	Mouse candidate	1
HEY1	rs1046472	Mouse candidate	1
HIST1H1C	rs10425	Mouse candidate	1
HK2	rs3821305	Mouse candidate	1
HLA-DPA1	rs7905	Mouse candidate	1
IGF2	rs680	Imprinted control	1
IGFBP1	rs4619	Human candidate	1
IL15	rs1057972	Mouse candidate	1
IL1RN	rs315951	Mouse candidate	1
ILK	rs1043388	Human candidate	1
IMPACT	rs1053474	Biallelic control	1
INPP5F	rs1063224	Biallelic control	1
IRS2	rs4773092	Mouse candidate	1
KHK	rs1131375	Mouse candidate	1
KIAA0523	rs3744725	Mouse candidate	1
KIAA1571	rs7582864	Mouse candidate	1
KRT6E	rs2568	Human and mouse candidate	1
LASS4	rs36260	Mouse candidate	1
LCP1	rs11342	Mouse candidate	1

LEMD2	rs2296744	Mouse candidate	1
MBP	rs9199	Human candidate	1
MFGE8	rs10859	Human candidate	failed
MRPS34	rs1076695	Human candidate	not expressed
MTMR3	rs41171	Mouse candidate	1
MYH7B	rs2425009	Mouse candidate	1
NEDD9	rs1050775	Mouse candidate	1
NR3C2	rs5534	Mouse candidate	not expressed
NUDCD1	rs1548082	Mouse candidate	1
NXPH1	rs3779355	Mouse candidate	1
OSBP2	rs2301816	Mouse candidate	1
PEG10	rs13073	Imprinted control	1
PEG3	rs1860565	Imprinted control	1
PERLD1	rs2952151	Mouse candidate	1
PHF11	rs1046295	Human candidate	1
PHLDA2	rs13390	Imprinted control	1
PIK3R1	rs3756668	Mouse candidate	1
PLB1	rs2272387	Mouse candidate	1
PLCL1	rs1064213	Mouse candidate	1
POGK	rs10918585	Mouse candidate	1
PON2	rs6954345	Imprinted in mouse	1
PPP1CB	rs7475	Mouse candidate	1
PRDM8	rs12780	Mouse candidate	1
PRKAR2B	rs257376	Mouse candidate	1
PSMB6	rs3169950	Mouse candidate	1
PTGFR	rs899	Mouse candidate	failed
PTPRB	rs2465811	Mouse candidate	1
RAFTLIN	rs842424	Human candidate	1
RAPGEF5	rs3779069	Mouse candidate	failed
RARRES1	rs2307064	Mouse candidate	1
RASGRF1	rs11855231	Imprinted in mouse	1
RFXDC2	rs3803459	Mouse candidate	1
SACS	rs4143768	Mouse candidate	not expressed
SGCD	rs7724969	Mouse candidate	not expressed
SHRM	rs3733242	Mouse candidate	1
SILV	rs1052165	Human candidate	1
SLC22A2	rs694812	Imprinted in mouse	failed
SLC22A3	rs2076828	Imprinted in mouse	1
SLC27A2	rs1648348	Mouse candidate	failed
SLC2A1	rs2229682	Human candidate	1
SLC40A1	rs2304704	Human candidate	1
SMARCA3	rs2119342	Mouse candidate	1
SNX19	rs3751037	Mouse candidate	1
SPARCL1	rs9933	Human candidate	1
SRP14	rs16924528	Mouse candidate	1
SSNA1	rs3087779	Mouse candidate	1
ST8SIA4	rs1428439	Mouse candidate	1
STOX1	rs10509305	Biallelic control	1

TCF20	rs2070116	Mouse candidate	1
TF	rs8649	Human candidate	1
TGFBI	rs1054124	Human and mouse candidate	1
TRMT12	rs3812475	Mouse candidate	1
TSPAN4	rs7091	Other candidate	1
TSPYL4	rs2232472	Mouse candidate	1
TSSC4	rs1057769	Biallelic control	no hets
USP29	rs3764574	Imprinted in mouse	failed
WNT2	rs2024233	Mouse candidate	1
YOD1	rs2629665	Mouse candidate	1
ZF	rs7116195	Mouse candidate	1
ZNF346	rs251848	Mouse candidate	1

Table 2: List of SNPs and genes tested on Illumina platform

Illumina		Human imprinted	Orthologues of mouse imprinted	Other controls	Mouse candidate	Human candidate	Sood et al. placenta	Genes associated with higher or lower birth weight	Polycomb gene	Housekeeping gene
rs12214	CTSD				x	x				
rs1053796	KRT6E				x	x				
rs12102	SERPINB2				x	x				
rs1054124	TGFBI				x	x				
rs1554005	ACOXL				x					
rs12449580	AIPL1				x					
rs2073478	ALDH1B1				x					
rs3591	ATP9B				x					
rs6753785	BCL2L11				x					
rs7013	C10orf10				x					
rs1043583	C10orf9				x					
rs11245007	C10orf90				x					
rs10140007	C14orf100				x					

rs1048257	C14orf78				**x**					
rs2279269	C18orf4				**x**					
rs7146	C19orf6				**x**					
rs1539172	C9orf93				**x**					
rs10781499	CARD9				**x**					
rs754615	CAST				**x**			**x**		
rs11802875	CDCA1				**x**					
rs1563727	CDKAL1				**x**					
rs1057874	CNN2				**x**					
rs2271029	CNTNAP1				**x**					
rs1050351	COL18A1				**x**					
rs11074359	COQ7				**x**					
rs1219725	CPXM2				**x**					
rs17459	CRYZ				**x**					
rs2830239	CYYR1				**x**					
rs233112	DDAH1				**x**					
rs1053227	DDIT4L				**x**					
rs2574	DGKG				**x**					
rs11122324	DISC1				**x**					
rs2271108	DOCK5				**x**					
rs3133	DTNA				**x**					
rs5333	EDNRA				**x**					
rs6072352	EMILIN3				**x**					
rs12439907	FLJ12994				**x**					
rs1042772	FMO4				**x**					
rs2279990	FOSL2				**x**					
rs17783344	GCA				**x**					
rs2301230	GNAI3				**x**					
rs2774315	GNG4				**x**					
rs13106386	GPR125				**x**					
rs10828833	GPR158				**x**					
rs707176	GRIA1				**x**					
rs1036199	HAVCR2				**x**					
rs1128159	HCLS1				**x**					
rs1046472	HEY1				**x**					
rs10425	HIST1H1C				**x**					
rs3732299	HK2				**x**					
rs8807	HLA-DPA1				**x**					
rs1047985	HLA-DQA1				**x**					
NT-007592.14-23572982	HLA-DQA2				**x**					
rs1047033	ID4				**x**					
rs1057972	IL15				**x**					
rs315951	IL1RN				**x**					
rs1865434	IRS2				**x**					
rs3802197	KCNS2				**x**					
rs1870377	KDR				**x**					
rs1131375	KHK				**x**					
rs11156878	KIAA0391				**x**					

rs6848033	KIAA0922				**x**					
rs3813360	KIAA1913				**x**					
rs12891	LASS4				**x**					
rs11342	LCP1				**x**					
rs10947436	LEMD2				**x**					
rs1397548	LPHN3				**x**					
rs2572925	LY6D				**x**					
rs237025	MAP3K7IP2				**x**					
rs2279241	MRAS				**x**					
rs2278125	MRVI1				**x**					
rs3087538	MSX2				**x**					
rs12537	MTMR3				**x**					
rs2425009	MYH7B				**x**					
rs1044417	NEDD9				**x**					
rs2871	NR3C2				**x**					
rs1131339	NR4A3				**x**					
rs2676793	NXPH3				x					
rs2273888	OGFRL1				**x**					
rs1328970	OPN5				**x**					
rs954474	OR2T1				**x**					
rs1055091	ORMDL1				**x**					
rs2301816	OSBP2				**x**					
rs9376173	PDE7B				**x**					
rs2941504	PERLD1				**x**					
rs12375	PEX14				**x**					
rs1082	PHACTR2				**x**		**x**			
rs3213563	PIGC				**x**					
rs3756668	PIK3R1				**x**					
rs2528588	PKP4				**x**					
rs2272386	PLB1				**x**					
rs1064213	PLCL1				**x**					
rs10918585	POGK				**x**					
rs3190	PPP1CB				**x**					
rs1537406	PRDM16				**x**					
rs12780	PRDM8				**x**					
rs257376	PRKAR2B				**x**					
rs2294008	PSCA				**x**					
rs12074883	PTGFR				**x**					
rs3842803	PTGS1				**x**					
rs2303963	PTPRB				**x**					
rs10206850	Q9C0I4_HUMAN/ KIAA1679				**x**					
rs1872575	QTRTD1				**x**					
rs12592	RAPGEF5				**x**					
rs10276	RARRES1				**x**					
rs1061033	RCN1				**x**					
rs2737700	SACS				**x**					
rs284445	SGCD				**x**					
rs4803	SLB				**x**					

rs1110277	SLC23A2				**x**					
rs5569	SLC6A2				**x**					
rs2119342	SMARCA3				**x**					
rs11250064	SOX7				**x**					
rs16924528	SRP14				**x**					
rs6715729	TACR1				**x**					
rs7309	TANK				**x**					
rs2292971	TBCD				**x**					
rs2070116	TCF20				**x**					
rs324356	TFB1M				**x**					
rs1051388	TMEM30B				**x**					
rs1931895	TSPYL4				**x**					
rs12464787	TTN				**x**					
rs2024233	WNT2				**x**					
rs11980379	ZNFN1A1				**x**					
rs10888390	CTSS					**x**				
rs2011951	LOC51629/ CGI-69					**x**		**x**		
rs1633462	SF1					**x**		**x**		
rs1385129	SLC2A1					**x**				
rs11539983	SLC40A1					**x**	**x**			
rs8177232	TF					**x**				
rs16943991	ABC1									
rs3744376	ABC1									
rs1128503	ABCB1									
rs2214102	ABCB1									
rs3742801	ABCD4									
rs4148077	ABCD4									
rs6925	ABHD6									
rs8723	ABHD6									
rs8176742	ABO									
rs8176749	ABO									
rs13475	ACN9									
rs7837	ACN9									
rs12985	ACO1									
rs3780473	ACO1									
rs137831	ACO2									
rs203316	ACO2									
rs7558938	ACOXL									
rs2245231	ACTR5									
rs1046003	ACY1L2									
rs1048641	ACY1L2									
rs6976	AD-017									
rs3743598	ADAT1									
rs2230739	ADCY9									
rs879620	ADCY9									
rs4698	ADM									
rs5005	ADM									
rs761745	ADM2									

rs1801253	ADRB1									
rs1042718	ADRB2									
rs1042719	ADRB2									
rs3087609	ADSS									
rs2269475	AIF1									
rs2736182	AIF1									
rs2292546	AIPL1									
rs2108978	AKAP10									
rs1130738	ALDH1A3									
rs3803430	ALDH1A3									
rs3043	ALDH1B1									
rs3741041	AMPD3			**x**						
rs2916747	ANGPT2									
rs3020221	ANGPT2									
rs1044250	ANGPTL4									
rs504574	ANK1									
rs750625	ANK1									
rs7816734	ANK1									
rs13309	AP4M1									
rs1534310	AP4M1									
rs7260921	APBA2BP									
rs1271	APOA1BP									
rs7412	APOE							**x**		
rs132653	APOL3									
rs707921	APOM									
rs10826997	ARHGAP12									
rs2070097	ARHGAP4									
rs766894	ARHGEF12									
rs9625	ARHGEF12									
rs10305751	ARNT									
rs11552229	ARNT									
rs2228099	ARNT									
rs4459508	ARNT2									
rs6495511	ARNT2									
rs7172548	ARNT2									
rs7484	ARNT2									
rs2071421	ARSA									
rs26653	ARTS-1									
rs27044	ARTS-1									
rs4859571	ASAHL									
rs16927574	ASPH									
rs6549	ASPH									
rs4281490	ASRGL1									
rs652313	ASS									
rs2295764	ASXL1								**x**	
rs4911231	ASXL1									
rs10475	ATF3									
rs283525	ATF5									

rs2066710	ATP10A	**x**								
rs3816800	ATP10A									
rs12154	ATPIF1									
rs9508	ATPIF1									
rs393521	AXIN1									
rs394128	AXIN1									
rs1041073	B3GTL									
rs876540	B3GTL									
rs3764779	B4GALT4									
rs1057077	B7									
rs710415	B7									
rs1055388	BAT1									
rs11796	BAT1									
rs1046080	BAT2									
rs1046089	BAT2									
rs10573	BAT5									
rs1475865	BAT5									
rs11539585	BBP									
rs17122715	BBP									
rs724710	BCL2L11									
rs4988398	BIK									
rs2306234	BLK									
rs3816668	BLK									
rs12250221	BLOC1S2									
rs4550	BNIP3			**x**						
rs6557	BNIP3									
rs12165	BNIP3L									
rs284854	C10orf26									
rs284860	C10orf26									
rs3812676	C10orf90									
rs7560	C14orf100									
rs2166322	C14orf118									
rs1043831	C14orf124									
rs10144530	C14orf130									
rs2905	C14orf130									
rs2275591	C14orf160									
rs3742935	C14orf78									
rs1061435	C16orf35									
rs2541622	C16orf35									
rs7665	C16orf5									
rs11151371	C18orf4									
rs699245	C1orf16									
rs840385	C1QTNF3									
rs840386	C1QTNF3									
rs2143607	C20orf111									
rs9875	C20orf111									
rs6107027	C20orf22									
rs7020	C20orf22									

rs3088196	C20orf44									
rs4911494	C20orf44									
rs3088078	C20orf52									
rs2056844	C21orf107									
rs6517523	C21orf107									
rs3746866	C21orf5									
rs1047978	C21orf91									
rs2824495	C21orf91									
rs1065201	C22orf2									
rs3747174	C22orf2									
rs1803196	C2orf3									
rs6722682	C2orf3									
rs2230204	C3									
rs423490	C3									
rs2501968	C6orf139									
rs130067	C6orf18									
rs3130453	C6orf18									
rs225710	C6orf55							**x**		
rs1620075	C8A									
rs652785	C8A									
rs20574	C8G									
rs2071006	C8G									
rs2409764	C8orf13									
rs3021518	C8orf13									
rs4741510	C9orf93									
rs443563	C9orf93									
rs1045882	C9orf95									
rs3752955	C9orf95									
rs1043239	CA12									
rs1043256	CA12									
rs12553173	CA9									
rs1801197	CALCR		**x**							
rs2301680	CALCR									
rs1043550	CALU									
rs8597	CALU									
rs1238	CAPZA1									
rs3135499	CARD15									
rs2075820	CARD4									
rs2043211	CARD8									
rs3745718	CARD8									
rs1135314	CARD9									
rs9667	CAST									
rs1049982	CAT									
rs8042868	CATSPER2									
rs10278782	CAV2									
rs1052990	CAV2									
rs1132644	CCNB1IP1									
rs1051130	CCND3							**x**		

rs9529	CCND3									
rs3136665	CCR1									
rs2228428	CCR4									
rs10946217	CCR6									
rs2071171	CCR6									
rs3093007	CCR6									
rs2229095	CCR7									
rs2853699	CCR8									
rs2012645	CCT8									
rs8129954	CCT8									
rs1130663	CD151									
rs1130719	CD151									
rs3211938	CD36									
rs1055141	CD4							**x**		
rs3829972	CD4									
rs11033026	CD44									
rs8193	CD44									
rs11585	CD59									
rs704697	CD59									
rs2070776	CD79B									
rs7921	CD79B									
rs1050650	CD83									
rs16874698	CD83									
rs1509022	CDCA1									
rs28216	CDH11							**x**		
rs35213	CDH11									
rs6633	CDK2AP1									
rs2501727	CDK5RAP2									
rs4836822	CDK5RAP2									
rs9465994	CDKAL1									
rs3217992	CDKN2B									
rs1594	CFLAR									
rs7573256	CFLAR									
rs1042180	CFTR									**x**
rs1800136	CFTR									
rs1127827	CGI-111									
rs6871	CGI-111									
rs1127149	CGI-49									
rs1054283	CGI-62									
rs13504	CGI-96									
rs1812240	CGI-96									
rs8136009	CGI-96									
rs816407	CHCHD2									
rs8406	CHCHD2									
rs7542034	CHI3L2									
rs8535	CHI3L2									
rs1045861	CHORDC1									
rs1053754	CHRNE									

rs8834	CHRNE									
rs2230804	CHUK									
rs1064108	CHURC1									
rs1131431	CITED2									
rs7313141	CLECSF2									
rs2272592	CLIC1									
rs3237	CLIC1									
rs1800209	CLN5									
rs1548082	CML66									
rs1548083	CML66									
rs1057895	CNN2									
rs2236451	COL18A1									
rs1042917	COL6A2									
rs2839110	COL6A2									
rs9843784	COMMD2									
rs11227	COPS7A									
rs3168600	COPS7A									
rs2275710	CORO2A									
rs701753	CP									
rs10492785	CP110									
rs12934510	CP110									
rs7190666	CP110									
rs2171492	CPA4	**x**								
rs1564823	CPEB4									
rs359467	CPEB4									
rs6060539	CPNE1									
rs6579255	CPNE1									
rs8277	CPXM2									
rs17047660	CR1									
rs2296160	CR1									
rs6691117	CR1									
rs3087822	CRIPT									
rs2255255	CRNKL1									
rs2255258	CRNKL1									
rs2273058	CRNKL1									
rs3817995	CRNKL1									
rs8140949	CRYBB2									
rs7527057	CRYZ									
rs1058885	CSF1									
rs3738760	CSF1									
rs216123	CSF1R						**x**			
rs2228422	CSF1R									
rs1042658	CSF3									
rs2827	CSF3									
rs3917991	CSF3R									
rs6385	CSTB									
rs3736213	CSTF3									
rs3758741	CSTF3									

rs1045480	CTBP1								**x**	
rs1048682	CTNS									
rs222754	CTNS									
rs8839	CTSD									
rs1036938	CTSH									
rs3129	CTSH									
rs17479770	CUL3								**x**	
rs4674908	CUL3									
rs630693	CWF19L2									
rs630782	CWF19L2									
rs3732379	CX3CR1									
rs7636125	CX3CR1									
rs2234355	CXCR6									
rs2234358	CXCR6									
rs13397	CXorf12									
rs6571303	CXorf12									
rs1048943	CYP1A1									
rs2470890	CYP1A2									
rs1056836	CYP1B1									
rs10916	CYP1B1									
rs451652	CYP21A2									
rs7756934	CYP21A2									
rs1137115	CYP2A6									
rs4986892	CYP2A6									
rs1042194	CYP2C18									
rs2860840	CYP2C18									
rs3758580	CYP2C19									
rs4244285	CYP2C19									
rs1058932	CYP2C8									
rs1057910	CYP2C9									
rs1799853	CYP2C9									
rs9332242	CYP2C9									
rs966410	CYYR1									
rs3761936	DCLRE1B									
rs6674384	DCLRE1B									
rs7277	DCTD									
rs7663494	DCTD									
rs233113	DDAH1									
rs805304	DDAH2									
rs4647707	DDB2									
rs4898778	DDHD1									
rs1053639	DDIT4									
rs8316	DDIT4									
rs11734833	DDIT4L									
rs1043402	DDX17									
rs763121	DDX17									
rs197414	DDX20									
rs85276	DDX20									

rs6444109	DGKG									
rs10186730	DGUOK									
rs6737156	DGUOK									
rs1044482	DHCR7			**x**						
rs1064202	DHX34									
rs2547378	DHX34									
rs2255397	DIP2A									
rs1411771	DISC1									
rs664628	DKFZP434B172									
rs6763762	DKFZP434B172									
rs1972576	DKFZP434F0318									
rs2110597	DKFZP434F0318									
rs384403	DKFZp434N035									
rs434049	DKFZp434N035									
rs7230131	DKFZP564D1378									
rs6559	DKFZP566H073									
rs1065584	DKFZP566J2046									
rs3743853	DKFZP566J2046									
rs2291617	DKFZP586D0919									
rs923829	DKFZP586D0919									
rs6660019	DKFZp761A078									
rs12582	DKFZp762E1312									
rs1275391	DLGAP4							**x**		
rs220079	DLGAP4									
rs1802710	DLK1	**x**								
rs17145034	DMXL1									
rs7734532	DMXL1									
rs1801041	DNA2L									
rs3758626	DNA2L									
rs11617079	DNAJD1									
rs3783044	DNAJD1									
rs2709618	DOCK5									
rs1052556	DPYSL4									
rs2247705	DPYSL4									
rs4764794	DRIM									
rs703715	DRIM									
rs7337	DSCR5									
rs9944927	DTNA									
rs7583475	DTNB									
rs11919795	DVL3							**x**		
rs5369	EDN1									
rs5370	EDN1									
rs5335	EDNRA									
rs2153364	EGLN1									
rs7544596	EGLN1									
rs2545763	EGLN2									
rs1680709	EGLN3									
rs1680710	EGLN3									

rs17085249	ELL2							**x**		
rs373533	EMR1									
rs1061223	ENO2									
rs799265	ENSG00000135506									
rs1230358	ENSG00000164308									
rs2549782	ENSG00000164308									
rs3752277	ENST00000360896									
rs10495933	EPAS1						**x**			
rs1868091	EPAS1									
rs126013	EPB41									
rs2249138	EPB41									
rs1042168	EPB42									
rs16957499	EPB42									
rs1051741	EPHX1									
rs2234922	EPHX1									
rs564449	EPO									
rs13181	ERCC2							**x**		
rs1799793	ERCC2									
rs12124733	ERMAP									
rs12727498	ERMAP									
rs13036061	ETAA16									
rs5960	F10									
rs4525	F5									
rs6030	F5									
rs6042	F7									
rs6046	F7									
rs440051	F9									
rs6048	F9									
rs1061646	FANCA									
rs7195066	FANCA									
rs8328	FBXO28									
rs1035834	FBXO36									
rs1801274	FCGR2A									
rs387801	FCGR2A									
rs844	FCGR2B						**x**			
rs396991	FCGR3B									
rs448740	FCGR3B									
rs2044174	FDX1									
rs11712	FKBP1A									
rs6041749	FKBP1A									
rs3824250	FLJ10204									
rs6470147	FLJ10204									
rs6999234	FLJ10204									
rs7014678	FLJ10204									
rs2788478	FLJ10300									
rs1551528	FLJ10305									
rs16970545	FLJ10305									
rs17075612	FLJ10375									

rs3772165	FLJ10375									
rs2242471	FLJ10498									
rs6830514	FLJ10525									
rs1876268	FLJ10858									
rs366793	FLJ10891									
rs433377	FLJ10891									
rs835409	FLJ10986									
rs835435	FLJ10986									
rs1561736	FLJ11184									
rs215210	FLJ11730									
rs9787162	FLJ11838									
rs1057090	FLJ12847									
rs3803459	FLJ12994									
rs3829533	FLJ12998									
rs3829536	FLJ12998									
rs2292071	FLJ13119									
rs745960	FLJ14640									
rs9409550	FLJ14753									
rs6962151	FLJ20257									
rs2015240	FLJ20444									
rs2307055	FLJ20444									
rs12142199	FLJ20542									
rs7627	FLJ20920									
rs9674937	FLJ20920									
rs3203	FLJ21945									
rs3731620	FLJ21945									
rs10216063	FLJ22374									
rs16875355	FLJ22374									
rs3739435	FLJ22494									
rs12459634	FLJ22573									
rs12801980	FLJ22635									
rs7104019	FLJ22635									
rs332259	FLJ22875									
rs6672905	FLJ32112									
rs7524477	FLJ32112									
rs2719710	FLJ32871			FLJ33071						
rs6599309	FLJ34443									
rs12912744	FLJ35867									
rs3809482	FLJ35867									
rs612448	FLJ37970									
rs685870	FLJ37970									
rs10205	FLOT2									
rs1060247	FLOT2									
rs7326277	FLT1									
rs7993418	FLT1									
rs11120047	FLVCR									
rs1155779	FMNL2							**x**		
rs4664114	FMNL2									

rs7101	FOS						**x**			
rs1049698	FOSB									
rs708905	FOSB									
rs7562	FOSL2									
rs7144658	FOXA1									
rs1044959	FOXO3A									
rs881732	FOXP2									
rs1042229	FPR1							**x**		
rs2070745	FPR1									
rs509474	FTHFSDC1									
rs7543	FTHFSDC1									
rs2584625	FTSJ3									
rs2727288	FTSJ3									
rs4015	FUT1						**x**			
rs4021	FUT1									
rs16880852	FUT10									
rs2676415	FUT10									
rs281377	FUT2									
rs485073	FUT2									
rs1050828	G6PD									
rs7320583	GAS6									
rs8191973	GAS6									
rs1058240	GATA3									
rs1049508	GATM		**x**							
rs1145086	GATM									
rs10493821	GBP3									
rs1409150	GBP3									
rs17433780	GBP3									
rs3795543	GBP3									
rs2592551	GGCX									
rs699664	GGCX									
rs1050160	GLUD1									
rs3737182	GNAI3									
rs1800900	GNAS	**x**								
rs3730171	GNAS									
rs8386	GNAS									
rs5446	GNB3									
rs2774316	GNG4									
rs6993	GOT2									
rs7202491	GOT2									
rs1048126	GPATC2									
rs4147127	GPATC2									
rs1864139	GPI									
rs1798192	GPR109B									
rs9002	GPR125									
rs2480345	GPR158									
rs9320308	GPR63									
rs1800504	GRB10	**x**								

rs3809806	GSG2									
rs33657	GSPT1							**x**		
rs3752426	GSPT1									
rs592792	GSTM2									
rs625456	GSTM2									
rs4630	GSTT1									
rs6488889	GTF2H3									
rs9530	GUSB									**x**
rs7658293	GYPA									
rs1849119	GYPB									
rs7683365	GYPB									
rs1050967	GYPC									
rs6568	GYPC									
rs1786702	H17									
rs594318	H17									
rs2075745	H19	**x**								
rs2839702	H19									
rs1057687	HABP4									
rs7030316	HABP4									
rs16942414	HAPLN3									
rs8039131	HAPLN3									
rs4704846	HAVCR2									
rs10780755	HBLD2							**x**		
rs7021024	HBLD2									
rs730106	HCFC1									
rs2070180	HCLS1									
rs2653349	HCRTR2									
rs1053657	HEBP2									
rs2232248	HEMK1									
rs667894	HEMK1									
rs1133496	HERC2									
rs4778244	HERC2									
rs3734637	HEY2									
rs5745635	HGF									
rs1058180	HIBCH									
rs291466	HIBCH									
rs17099141	HIF1A									
rs2057482	HIF1A									
rs10883512	HIF1AN									
rs2295778	HIF1AN									
rs4803932	HIF3A									
rs7253301	HIF3A									
rs8384	HIST1H1C									
rs656489	HK2									
rs1049281	HLA-C							**x**		
rs1094	HLA-C									
rs7767581	HLA-C									
rs11244	HLA-DOB									

rs2071469	HLA-DOB									
rs9272934	HLA-DQA1									
rs9272953	HLA-DQA1									
rs2071799	HLA-DQA2									
rs9276401	HLA-DQA2									
rs3211055	hmm25128									
rs3752108	hmm25128									
rs3811073	hmm31999									
rs4840040	hmm31999									
rs4962697	hmm665									
rs11555832	HMOX1									
rs2071747	HMOX1									
rs1051308	HMOX2									
rs17137094	HMOX2									
rs4885028	Hs.28465									
rs9543091	Hs.28465									
rs476240	Hs.345389									
rs6423498	Hs.345389									
rs1875428	Hs.406038									
rs2291782	Hs.506072									
rs7725810	Hs.506072									
rs17137053	HSCARG									
rs17137056	HSCARG									
rs1061810	HSD17B12									
rs162398	HSF2BP									
rs2838343	HSF2BP									
rs1043618	HSPA1A									
rs2075799	HSPA1L									
rs2075800	HSPA1L									
rs16833517	HSPBAP1									
rs1059384	HSRTSBETA									
rs2612086	HSRTSBETA									
rs5490	ICAM1							**x**		
rs1799969	ICAM1									
rs5491	ICAM1									
rs5498	ICAM1									
rs303172	IFIT5									
rs304447	IFIT5									
rs11914	IFNGR1									
rs1327475	IFNGR1									
rs1059293	IFNGR2									
rs11910627	IFNGR2									
rs9808753	IFNGR2									
rs1003483	IGF2AS	**x**								
rs3741208	IGF2AS									
rs1570070	IGF2R									
rs998075	IGF2R									
rs1059713	IGHM									

rs3024496	IL10									
rs3024498	IL10									
rs2256111	IL10RA									
rs9610	IL10RA									
rs1058867	IL10RB						**x**			
rs3171425	IL10RB									
rs3212227	IL12B									
rs436857	IL12RB1									
rs2229546	IL12RB2									
rs20541	IL13									
rs1519553	IL15									
rs4072680	IL16									
rs859	IL16									
rs3017	IL17RB									
rs6798958	IL17RB									
rs360717	IL18									
rs549908	IL18									
rs1035130	IL18R1									
rs3732127	IL18R1									
rs1304037	IL1A									
rs2856836	IL1A									
rs1143634	IL1B									
rs315952	IL1RN									
rs2069763	IL2									
rs10889677	IL23R									
rs1884444	IL23R									
rs10903034	IL28RA									
rs11249006	IL28RA									
rs1805011	IL4R									
rs8832	IL4R									
rs2290610	IL5RA									
rs340833	IL5RA									
rs2069849	IL6									
rs2229238	IL6R									
rs4845617	IL6R									
rs1111046	IMAGE3451454									
rs2236359	IMAGE3451454									
rs6750289	IMMT									
rs8244	IMMT									
rs1053474	IMPACT									
rs582234	IMPACT									
rs1475563	INADL									
rs2498982	INADL									
rs4764616	ING4									
rs2289306	INPP5F									
rs3188055	INPP5F									
rs6908105	IRAK1BP1									
rs1152888	IRAK3									

rs7135413	IRAK3									
rs13180	IREB2									
rs3743079	IREB2									
rs839	IRF1									
rs9282762	IRF1									
rs3775543	IRF2									
rs2304204	IRF3									
rs1050975	IRF4									
rs7768807	IRF4									
rs2013162	IRF6									
rs3178010	IRF7									
rs4773092	IRS2									
rs3809865	ITGB3									
rs5919	ITGB3									
rs7004	ITGB4BP									
rs10425594	IXL									
rs519575	IXL									
rs17127063	JAK1									
rs2230724	JAK2									
rs3008	JAK3									
rs8234	KCNQ1	**x**								
rs7667298	KDR									
rs2304681	KHK									
rs2813	KIAA0089									
rs6799559	KIAA0089									
rs6794	KIAA0116									
rs1121	KIAA0251									
rs6498540	KIAA0251									
rs12534379	KIAA0265									
rs3742370	KIAA0391									
rs17135121	KIAA0415									
rs9790	KIAA0643									
rs2252690	KIAA0748									
rs2252795	KIAA0748									
rs7669418	KIAA0922									
rs16989000	KIAA1271									
rs3746660	KIAA1271									
rs7262903	KIAA1271									
rs7269320	KIAA1271									
rs659543	KIAA1324									
rs1064034	KIAA1627									
rs4834698	KIAA1627									
rs1553669	KIAA1712									
rs4695918	KIAA1712									
rs3813359	KIAA1913									
rs2236599	KLF4									
rs2290019	KLHDC4									
rs2303771	KLHDC4									

rs583678	KRT6E									
rs3751325	KUB3									
rs2071970	L3MBTL	**x**						**x**		
rs2285185	L3MBTL									
rs36259	LASS4									
rs7188975	LCMT1									
rs1409429	LCP1									
rs4820	LDHA									
rs1650294	LDHB									
rs1433099	LDLR									
rs5927	LDLR									
rs2296744	LEMD2									
rs8368	LGALS9							**x**		
rs3745871	LILRB4							**x**		
rs731170	LILRB4									
rs1107853	LNK									
rs739496	LNK									
rs7972796	LNK									
rs1063677	LOC114984									
rs7281	LOC114984									
rs3740957	LOC119710									
rs1757935	LOC132321									
rs337277	LOC132321									
rs1054174	LOC133957									
rs13474	LOC133957									
rs1052202	LOC144404									
rs1052204	LOC144404									
rs12984381	LOC147804									
rs1818989	LOC148213									
rs1061860	LOC151963									
rs7507	LOC151963									
rs2014220	LOC200933									
rs3734502	LOC282956									
rs6904200	LOC282956									
rs16973457	LOC283726									
rs7403244	LOC283726									
rs2535241	LOC346171									
rs2747421	LOC346171									
rs12442603	LOC348094									
rs2414865	LOC348094									
rs7101779	LOC374421									
rs7131178	LOC374421									
rs1578462	LOC375097									
rs3752278	LOC375097									
rs17635222	LOC375399									
rs7697056	LOC375399									
rs2918520	LOC401075									
rs2436487	LOC56931									

rs1133122	LOC90637									
rs17273079	LOC93349									
rs7590429	LOC93349									
rs734644	LPHN3									
rs2254522	LSS									
rs1052248	LST1									
rs1041981	LTA									
rs2239704	LTA									
rs3093553	LTB									
rs13295	LY6G5C									
rs9328374	LY86									
rs7450	MAK10									
rs1059442	MALT1									
rs2319974	MALT1									
rs3797762	MAML1									
rs6627	MAML1									
rs1146297	MAN1A2									
rs1290558	MAN1A2									
rs1128933	MAN2C1							**x**		
rs10250	MAP2K2									
rs3732209	MAP3K2									
rs7896	MAP3K7IP2									
rs2907	MAP3K8									
rs8177039	MAP3K8									
rs7153601	MAP3K9									
rs6544214	MAP4K3									
rs9037	MAP4K3									
rs13058	MAPK1									
rs10764686	MASTL									
rs1981296	MASTL									
rs2099903	MBL2									
rs2506	MBL2									
rs2295709	MCMDC1									
rs2734647	MECP2									
rs12140829	MED8									
rs839753	MED8									
rs10863	MEST	**x**								
rs41736	MET									
rs3741265	MGC:13379									
rs8064449	MGC10744									
rs8069739	MGC10744									
rs1047707	MGC12458									
rs10927387	MGC12458									
rs842259	MGC15763									
rs842274	MGC15763									
rs7226091	MGC16597									
rs1055636	MGC19764									
rs3803859	MGC19764									

rs10152546	MGC20481									
rs1044474	MGC20481									
rs1046404	MGC20781									
rs4796712	MGC20781									
rs11806946	MGC22773									
rs1412825	MGC22773									
rs17361819	MGC22960									
rs3737744	MGC22960									
rs1061128	MGC24665									
rs7204628	MGC24665									
rs690941	MGC2744									
rs17194861	MGC2747									
rs706762	MGC2747									
rs10448	MGC2752									
rs3499	MGC2752									
rs3735169	MGC3036									
rs227584	MGC3130									
rs1339374	MGC31967									
rs3852768	MGC3248									
rs8056871	MGC3248									
rs353255	MGC3265									
rs2257505	MGC33648									
rs652541	MGC33948									
rs664143	MGC33948									
rs6585	MGC5242									
rs10489177	MGC9084									
rs1063635	MICA									
rs1131896	MICA									
rs1051788	MICB									
rs3131639	MICB									
rs1044483	MK-STYX									
rs8565	MK-STYX									
rs2287074	MMP2									
rs243849	MMP2									
rs6060341	MMP24									
rs1571133	MMRP19									
rs1977420	MMRP19									
rs2956114	MMRP19									
rs594445	MOCOS									
rs11872520	MPPE1									
rs3747956	MR1									
rs3789357	MR1									
rs1802752	MRPL28									
rs3830160	MRPL28									
rs2863095	MRPL43									
rs3740484	MRPL43									
rs1047911	MRPL53									
rs13960	MRPL54									

rs2070250	MRPS12									
rs17634737	MRPS9									
rs4909945	MRVI1									
rs4242182	MSX2									
rs10266424	MTERF									
rs9008	MTERF									
rs41172	MTMR3									
rs12347	MTRR									
rs162036	MTRR									
rs8659	MTRR									
rs6853	MYD88									
rs2425012	MYH7B									
rs3407	MYOM2									
rs968381	MYOM2									
rs7188856	NAGPA									
rs3829567	NARF									
rs16923269	NCBP1							**x**		
rs14189	NCOA3							**x**		
rs11699879	NCOA3						**x**			
rs1537028	NCOA5									
rs17092079	NCOA6									
rs1422645	NDP52									
rs2303015	NDP52									
rs7222365	NDP52									
rs1899	NDUFAF1									
rs3204853	NDUFAF1									
rs4148973	NDUFV3									
rs2303579	NEDD4									
rs1050775	NEDD9									
rs1957106	NFKBIA									
rs696	NFKBIA									
rs8904	NFKBIA									
rs7116	NM_001535									
rs2071128	NM_003491									
rs2070426	NM_006031									
rs2249057	NM_006031									
rs2241384	NM_012276									
rs7256494	NM_012276									
rs702681	NM_152622									
rs3848713	NM_183386									
rs6753	NOL5A							**x**		
rs1060037	NOL6									
rs2785210	NOL6									
rs1047735	NOS1									
rs2682826	NOS1									
rs3741475	NOS1									
rs10459953	NOS2A									
rs1137933	NOS2A									

rs2297518	NOS2A									
rs1799983	NOS3									
rs3918211	NOS3									
rs7830	NOS3									
rs2229974	NOTCH1									
rs4489420	NOTCH1									
rs6563	NOTCH1									
rs699779	NOTCH2									
rs1044116	NOTCH3									
rs16980398	NOTCH3									
rs422951	NOTCH4									
rs915894	NOTCH4									
rs3745615	NPAS1									
rs1562313	NPAS2									
rs2305158	NPAS2									
rs6026468	NPEPL1									
rs1800566	NQO1									
rs1054190	NR1I2									
rs3814057	NR1I2									
rs2307424	NR1I3									
rs5534	NR3C2									
rs2900223	NR4A3									
rs2229741	NRIP1									
rs2894215	NUDT13									
rs7025269	NUDT2									
rs7039222	NUDT2									
rs7124513	NUDT8									
rs11937770	NUDT9									
rs132848	NUP50									
rs1051042	OAS1									
rs2660	OAS1									
rs3741981	OAS1									
rs1058128	OGDH									
rs14239	OGDH									
rs11101224	OGDHL									
rs13736	OGDHL									
rs954475	OR2T1									
rs3182034	ORM1									
rs2289404	ORMDL1									
rs3804085	OSBP2									
rs701265	P2RY1									
rs8380	P4HA1									
rs1127091	P66beta									
rs9426938	P66beta									
rs7074	PACSIN2							**x**		
rs8569	PACSIN2									
rs1138800	PANX1									
rs2240542	PASK									

rs6709462	PASK									
rs7058209	PDHA1									
rs709610	PDHA1									
rs400037	PDIA2									
rs1063647	PDK2									
rs3178055	PDK2									
rs3064	PDLIM2									
rs3735893	PDLIM2									
rs3746954	PDXK									
rs2812	PECAM1									
rs6809	PECAM1									
rs9908930	PECAM1									
rs13073	PEG10	**x**								
rs1055359	PEG3	**x**								
rs1860565	PEG3									
rs2952151	PERLD1									
rs1057225	PEX5									
rs9462859	PEX6									
rs1064891	PFKFB3									
rs1539232	PFKFB3									
rs1057034	PFKL									
rs1049392	PFKM									
rs8716	PFKM									
rs2279211	PFKP									
rs11121567	PGD									
rs12942703	PGS1									
rs2292642	PGS1									
rs2074038	PHACS									
rs3107275	PHACS									
rs2073214	PHACTR2									
rs1049925	PHC1								**x**	
rs11061	PHC2								**x**	
rs5861	PHC2									
rs6425816	PHC2									
rs734094	PHEMX									
rs1056567	PHF19								**x**	
rs1837	PHF19									
rs13390	PHLDA2	**x**								
rs11133	PHYH									
rs473407	PHYH									
rs1050057	PIAS3									
rs2289863	PIAS4									
rs2289865	PIAS4									
rs2230471	PIGC									
rs11306	PIGF									
rs1028307	PIGT									
rs707577	PIGT									
rs706713	PIK3R1									

rs16865250	PIP5K1A									
rs2074957	PIP5K1C									
rs4807493	PIP5K1C									
rs1053454	PIP5K2A									
rs10828317	PIP5K2A									
rs1673407	PKHD1L1									
rs1673408	PKHD1L1									
rs16879659	PKHD1L1									
rs1783147	PKHD1L1									
rs1783174	PKHD1L1									
rs4735133	PKHD1L1									
rs1052176	PKLR									
rs932972	PKLR									
rs1042728	PKN1									
rs4926219	PKN1									
rs1049846	PLAGL1	**x**								
rs9373409	PLAGL1									
rs7601771	PLB1									
rs753381	PLCG1									
rs2228135	PLCL1									
rs1061307	PLSCR1									
rs1130809	PNRC1									
rs12445	PNRC1									
rs3820387	POGK									
rs1561328	POLR1A									
rs2288120	POLR1A									
rs6843	POLR2E									
rs1131383	POLR2J									
rs6591	POLR2L									**x**
rs1055177	POLR3F									
rs1474974	POLR3F									
rs7493	PON2		**x**							
rs2290417	PP3856									
rs896954	PP3856									
rs13787	PPA2									
rs4699179	PPA2									
rs546502	PPFIA1									
rs552282	PPFIA1									
rs7562391	PPIL3									
rs7475	PPP1CB									
rs2694657	PPP1R12A									
rs3203905	PPP1R12A									
rs854524	PPP1R9A	**x**								
rs854541	PPP1R9A									
rs2236696	PRDM15									
rs4075967	PRDM15									
rs885821	PRF1									
rs257377	PRKAR2B									

rs1799810	PROC									
rs5937	PROC									
rs6123	PROS1									
rs2288920	PRRG2									
rs3745474	PRRG2									
rs2976396	PSCA									
rs1803415	PSMF1									
rs3087751	PSMF1									
rs1805155	PTCH									
rs1055340	PTER									
rs1331255	PTER									
rs899	PTGFR									
rs5788	PTGS1									
rs5275	PTGS2									
rs689470	PTGS2									
rs2465811	PTPRB									
rs1048821	Q9C0I4_HUMAN/ KIAA1679									
rs1026619	QRSL1									
rs2015205	QRSL1									
rs2302188	R29124_1									
rs714106	R29124_1									
rs11015859	RAB18									
rs12248740	RAB18									
rs1065544	RAB31						**x**			
rs557706	RAB31									
rs11105	RAB40B									
rs1131906	RABIF									
rs373572	RAD18									
rs11855560	RAD51									
rs1801320	RAD51									
rs2251660	RAD52B									
rs1057957	RANBP2							**x**		
rs3744806	RAPGEFL1									
rs7221536	RAPGEFL1									
rs1567962	raptor									
rs3751934	raptor									
rs2307064	RARRES1									
rs11855231	RASGRF1		**x**							
rs1562008	RASGRF1									
rs3759091	RBMS2									
rs941208	RBMS2									
rs6711	RCN1									
rs3138142	RDH5									
rs1044418	REPS1									
rs4606	RGS2									
rs2273782	RLN1									
rs7225888	RNF135									
rs1200345	RPAP1									

rs721772	RPAP1									
rs2280370	RPL13									
rs12981911	RPL28									
rs7255657	RPL28									
rs1054427	RPL36AL									
rs2985698	RPL36AL									
rs10274	RPS6KB2									
rs13859	RPS6KB2									
rs2296308	RWDD3									
rs1468542	SACM1L									
rs2271619	SACM1L									
rs4770433	SACS									
rs9403	SARS2									
rs2304210	SBP1									
rs16973530	SCAMP5									
rs7174129	SCAMP5									
rs838891	SCARB1									
rs838896	SCARB1									
rs8475	SCARB2									
rs3210400	SCLY									
rs6555055	SDHA									
rs6962	SDHA									
rs4612984	SEC10L1									
rs8003535	SEC10L1									
rs4786	SELE									
rs6128	SELP									
rs6133	SELP									
rs9874	SELS									
rs6493090	SERF2									
rs15286	SERPINB1							**x**		
rs386713	SERPINB1									
rs17072097	SERPINB10									
rs724558	SERPINB10									
rs8097425	SERPINB10									
rs963075	SERPINB10									
rs9967382	SERPINB10									
rs6098	SERPINB2									
rs5878	SERPINC1									
rs11178	SERPINE1									
rs7242	SERPINE1									
rs2289993	SESTD1									
rs523200	SF1									
rs12264	SFRS6									
rs7724969	SGCD									
rs1281149	SH3TC1									
rs922521	SH3TC1									
rs12952556	SHMT1									
rs1979276	SHMT1									

rs1979277	SHMT1									
rs16940043	SIAT8E									
rs3210908	SIGIRR									
rs1132975	SIVA									
rs149411	SLC11A2									
rs150909	SLC11A2									
rs202391	SLC13A3									
rs7169	SLC16A1									
rs1131633	SLC16A3									
rs3763980	SLC16A7									
rs367035	SLC22A18	**x**								
rs624249	SLC22A2		**x**							
rs316003	SLC22A2									
rs2076828	SLC22A3		**x**							
rs2292334	SLC22A3									
rs1131382	SLC23A2									
rs2301629	SLC25A13							**x**		
rs10998219	SLC25A16									
rs1136645	SLC25A16									
rs4658	SLC2A1									
rs1894822	SLC2A14									
rs7966601	SLC2A3									
rs15300	SLC35B3									
rs3757099	SLC35B3									
rs390840	SLC37A1									
rs454849	SLC37A1									
rs2304704	SLC40A1									
rs2072081	SLC4A1									
rs5810	SLC9A1									
rs1043292	SMAP-5									
rs4616931	SMAP-5									
rs3182285	SMARCA3									
rs1061063	SNAP29									
rs178077	SNAP29									
rs3744346	SNX11									
rs7222136	SNX11									
rs4969168	SOCS3									
rs1952085	SOCS4									
rs3768720	SOCS5									
rs4953419	SOCS5									
rs2231562	SOCS6									
rs2536512	SOD3									
rs2695232	SOD3									
rs3100132	SORD									
rs3741651	SP1									
rs7927406	SPA17									
rs3803680	SPG7									
rs8060502	SPG7									

rs12090314	SPTA1									
rs3737515	SPTA1									
rs3738791	SPTA1									
rs1626923	SPTB									
rs229586	SPTB									
rs7035964	SPTLC1									
rs10277	SQSTM1									
rs4797	SQSTM1									
rs11577179	SSR2									
rs4661079	SSR2									
rs10745330	ST7L									
rs3790611	ST7L									
rs2066811	STAT2									
rs1053005	STAT3									
rs3198502	STAT5A									
rs2230097	STAT5B									
rs4559	STAT6									
rs194520	STEAP2									
rs194524	STEAP2									
rs13013693	STK25									
rs2279845	STK25									
rs524492	SUPT3H									
rs9369514	SUPT3H									
rs12335	SURF6									
rs2491	SURF6									
rs1061844	SVH									
rs5757650	SYNGR1									
rs756640	SYNGR1									
rs3183175	SYNGR3									
rs7533455	SYTL1									
rs8533	SYTL1									
rs550404	SYTL2							**x**		
rs641393	SYTL2									
rs881	TACR1									
rs13501	TAP2									
rs241448	TAP2									
rs1059288	TAPBP									
rs2071888	TAPBP									
rs2041385	TAPBPL									
rs2532500	TAPBPL									
rs2297208	TBC1D4									
rs7327548	TBC1D4									
rs4986131	TBCD									
rs7502875	TBX21									
rs5758651	TCF20									
rs12190287	TCF21							**x**		
rs2249778	TCL6									
rs6517105	TCP10L									

rs12484449	TEF									
rs9611577	TEF									
rs8649	TF									
rs912722	TFB1M									
rs3817672	TFRC						**x**			**x**
rs406271	TFRC									
rs473698	TGFA									
rs503314	TGFA									
rs2241715	TGFB1									
rs900	TGFB2									
rs4669	TGFBI									
rs6356	TH		**x**							
rs6357	TH									
rs2126854	THAP6									
rs9307834	THAP6									
rs1478604	THBS1									
rs2292305	THBS1									
rs256996	TICAM2									
rs1043968	TINP1									
rs6874609	TINP1									
rs8177376	TIRAP									
rs2282336	TJP2									
rs2309428	TJP2									
rs4833095	TLR1									
rs5743596	TLR1									
rs10776483	TLR10									
rs11096957	TLR10									
rs4274855	TLR10									
rs3804099	TLR2									
rs3804100	TLR2									
rs3775290	TLR3									
rs3775291	TLR3									
rs5030710	TLR4									
rs7869402	TLR4									
rs3775073	TLR6									
rs3821985	TLR6									
rs352140	TLR9									
rs3196765	TMEM30B									
rs2834217	TMEM50B									
rs17136392	TMEM8									
rs2071915	TMEM8									
rs3093665	TNF									
rs1059501	TNFRSF7									
rs11301	TOMM20							**x**		
rs1804644	TOMM34									
rs2450772	TOP1MT									
rs1058378	TOP2B									
rs1042522	TP53									

rs1059857	TRAP1									
rs7200737	TRAP1									
rs11217125	TRAPPC4									
rs569	TRAPPC4									
rs1048705	TRIM4									
rs2247762	TRIM4									
rs3738671	TRIT1									
rs7315	TRIT1									
rs1705805	TRNT1									
rs334773	TRNT1									
rs11070795	TRPM7									
rs616256	TRPM7									
rs11688004	TSGA10									
rs1573413	TSGA2									
rs2839536	TSGA2									
rs6755	TSPYL4									
rs17682	TSSC1									
rs2835655	TTC3							**x**		
rs2154538	TTC3									
rs2244492	TTN									
rs280523	TYK2									
rs6554	UBA52									
rs2838677	UBE2G2									
rs760431	UBE2G2									
rs13392	UBE2L3									
rs7444	UBE2L3									
rs1049871	UBE2V1									
rs8585	UBE2V1									
rs10929303	UGT1A1									
rs8330	UGT1A1									
rs4694697	UGT2B11									
rs4348159	UGT2B7									
rs7439366	UGT2B7									
rs12726	UNC13B									
rs10901439	UROS									
rs2027515	UROS									
rs2292807	USMG5									
rs7911488	USMG5									
rs8103779	USP29		**x**							
rs1027392	USP29									
rs3088040	USP36									
rs1918496	USP52									
rs6583048	VAV3									
rs8458	VAV3									
rs3176879	VCAM1									
rs3783613	VCAM1									
rs11527434	VDAC2									
rs2010963	VEGF									

rs25648	VEGF									
rs1642742	VHL									
rs779805	VHL									
rs1049341	VIM									
rs165531	VIM									
rs2359612	VKORC1									
rs7294	VKORC1									
rs1177562	VPS11									
rs4614	VPS11									
rs10409482	VRK3									
rs1052498	VRK3									
rs11547883	VRK3									
rs11879620	VRK3									
rs1057990	WARS2									
rs2645294	WARS2									
rs3790549	WARS2									
rs13232463	WBSCR27									
rs11956837	WDR36									
rs15736	WDR4									
rs2248490	WDR4									
rs6586250	WDR4									
rs11247226	WINS1									
rs12157	WINS1									
rs2411837	WINS1									
rs8451	WINS1									
rs2228946	WNT2									
rs9303634	WSB1							**x**		
rs2288034	WWOX									
rs383362	WWOX									
rs1884725	XDH									
rs2295475	XDH									
rs4944960	XRRA1									
rs9444	XRRA1									
rs4931	YWHAB									
rs8356	YWHAB									
rs2291940	ZCCHC9									
rs1783978	ZDHHC5							**x**		
rs1783979	ZDHHC5									
rs2789	ZFP161									
rs990072	ZFP161									
rs4801433	ZIM3		**x**							
rs7251328	ZIM3									
rs1122955	ZNF132									
rs1465789	ZNF132									
rs10405102	ZNF160									
rs329733	ZNF160									
rs9302870	ZNF200									
rs11879465	ZNF211									

rs3746219	ZNF211									
rs9749449	ZNF211									
rs2735537	ZNF213									
rs12753	ZNF230									
rs2356549	ZNF230									
rs7988277	ZNF237									
rs9579717	ZNF237									
rs10414299	ZNF264		**x**							
rs917340	ZNF264									
rs10515	ZNF266									
rs1978713	ZNF266									
rs7975069	ZNF268									
rs1811	ZNF30									
rs765746	ZNF30									
rs12982082	ZNF331	**x**								
rs8100247	ZNF331									
rs10413068	ZNF473									
rs16981706	ZNF473									
rs12975981	ZNF493									
rs4461198	ZNF493									
rs10217154	ZNF510									
rs12347533	ZNF510									
rs966591	ZNF557									
rs12609890	ZNF587									
rs2270494	ZNF597									
rs37824	ZNF597									
rs10413287	ZNF626									
rs13292096	ZNF79									
rs4504745	ZNF79									
rs12972502	ZNF85									
rs1048412	ZNRD1									
rs7770557	ZNRD1									
rs17152433	ZRANB1									
rs1045493	ZSWIM3									
rs2903808	ZSWIM3									

Appendix 2: Primers for Sequenom platform

GENE	SNP_ID	2nd-PCRP	1st-PCRP	UEP_SEQ	Term	Ext Call
ACAS2	rs4911163	ACGTTGGATGTTACTG GAAGACTCCATGCC	ACGTTGGATGGGTAGT TGCTCCTTTCATCC	GTGACTAAAG GGAAAATCTT	ACG	C
ACOXL	rs7558938	ACGTTGGATGCGTCTG TTCTGTCAACTCTG	ACGTTGGATGACAAA ACAAAACACAAATAC	TTCGAAAACCT CCATGCCC	ACG	G
ADAR	rs1127326	ACGTTGGATGCTGCAA AGGCATCCTTAGTC	ACGTTGGATGCTGGTA GAGACATTCACATC	TCAAAAGTAC AGAGAGCA	ACG	G
ADPGK	rs9460	ACGTTGGATGCCAACA GGCTGGAATGTACC	ACGTTGGATGGGTTTG GAAGACAGCCAAAG	GCTGGAATGT ACCTGATACA	ACT	A
AGPAT5	rs2911970	ACGTTGGATGGCTCTG CATTTCTTCCTAGC	ACGTTGGATGTTGGAT CCCGTCAGTGGTG	TTTCTTCCTAG CAAGTCT	ACT	T
ALDH1B1	rs2073478	ACGTTGGATGGGATG GATGCCTCTGAGCG	ACGTTGGATGAAGTAG ACTCGATCCCGCTC	CGGGGCCGGC TGCTGAACC	ACT	T
AMPD3	rs3741041	ACGTTGGATGGAGGTT GTTGAGCACCATGA	ACGTTGGATGACCCAC CCTACAGCTACTAC	AGCACCATGA TGTTGGC	ACT	T
ANTXR2	rs7747	ACGTTGGATGTTCTCC CTTCCAATGTGATG	ACGTTGGATGTCTCTC TGTGAGCACAAAGG	GTGATGGTTGT TGAACTTAT	ACG	C
ATG16L	rs1045100	ACGTTGGATGGAGCTT TCTGGCTCTCTTTC	ACGTTGGATGCCATTT TTTAAAGCCACACGG	TCCCCCACAAA ATTCGACA	ACG	C
ARHGAP28	rs4239328	ACGTTGGATGACCGA ATGAGTCTGTGGAAC	ACGTTGGATGCAGAGT GTTTGCTTCTACTG	TTTCTACAGTG ATGGCACC	ACT	A
ATP10A	rs2066710	ACGTTGGATGAGCCTC TAAAAGGACTTGCC	ACGTTGGATGAGACA GTGTGTCCAGTTACC	CCTTACGGAA GTGCCCT	ACG	G
ATP9B	rs3591	ACGTTGGATGGACTGA AAGTTTAATTTCTGC	ACGTTGGATGTGAGTA GGAATGACTTTGTG	AATTTCTGCAG TTCCCTCA	ACG	C
BCL2L11	rs6753785	ACGTTGGATGAATCTC CTTGGGACTTTGAC	ACGTTGGATGATGGGC TCGAGACACATCAG	AGCTGTGTTGG AGCTCA	ACT	T
BMP6	rs17557	ACGTTGGATGGTGGCT TTCTTCAAAGTGAG	ACGTTGGATGATTACG ACTCTGTTGTCGGC	TTCAAAGTGA GTGAGGT	ACT	C
BSG	rs8259	ACGTTGGATGGCTCTC GTGAAACACTTCAG	ACGTTGGATGCTCCAT TCAGGATTCTGTTC	ACACTTCAGA AGGAAAAAA	CGT	T
C10orf9	rs1043583	ACGTTGGATGCATGGC GCAGACTTCCAAAT	ACGTTGGATGAGAAA GCTCAGAAGACAGCG	AGTACTGGTCA TTTCTCTC	ACG	C
C14orf100	rs7560	ACGTTGGATGTATCTG CTACACTGGAAGG	ACGTTGGATGACACAT TGCTAGGTACACAG	CGCTAGGAAG CCCTTGC	ACT	T
FLJ13154/ C16orf57	rs11551263	ACGTTGGATGACTCTG ATCTGAGGCCAATG	ACGTTGGATGTAGTCC CCCTAACAAACAGC	ACAAATTGCTG AAGTGG	ACT	T
C18orf4	rs7227616	ACGTTGGATGGCCAA AGATACTTGGTGATG	ACGTTGGATGAAAGA AGAACCTCCTCCACC	ACTTGGTGATG TACAGAAAT	ACG	C
MGC13170/ C19orf48	rs9991	ACGTTGGATGCATGAC ATCTGGCCAGTGAC	ACGTTGGATGCACACC CTGACACCATGTTC	AGACACCCTTC TCTTCC	ACT	A
C19orf6	rs7146	ACGTTGGATGACATCC ACATCGTCTTCTCC	ACGTTGGATGCACTTC CACACGCAGGATG	ATCAACTGCCT GGAGCA	ACG	G
C1orf121	rs2242448	ACGTTGGATGTTTGTG TGCCTATTTATGCC	ACGTTGGATGGTCTTC AAAAATTAGGCAGG	TTATGCCATAG ATCTGGCAA	ACG	G
C1orf164	rs1051664	ACGTTGGATGGACACC ACAGGAGAGAGAG	ACGTTGGATGTTTTAA GAACCCCTGGAAGC	AGGAGAGAGA GAGGCAG	ACT	T
C22orf25	rs737986	ACGTTGGATGCTGGCC GACTCCGCTAGTG	ACGTTGGATGACCCAC CTTCAAGCAGCCT	CTAGTGGCCCG GCCGGCCTG	ACT	T
C9orf93	rs1539172	ACGTTGGATGTTTGTC AGCTTCCTCACTGG	ACGTTGGATGACAGGC ACAACAACTACAGG	TGGAGTGCAA TTCAAGT	ACG	G
CALCR	rs1801197	ACGTTGGATGCTGGCG ACATCCCAATTTAC	ACGTTGGATGATGATC TCAGCACTCTCCTC	CATCTGCCATC AGGAGC	ACG	G
CARD9	rs10781499	ACGTTGGATGCACGCA TCACACCTTACCTG	ACGTTGGATGACTTTC CGTTTGCGGATGAC	CAGTGCAAGG TCCTGAACCC	ACG	G
ERC2_HUM AN/CAST	rs754615	ACGTTGGATGGGGTGC TGACTGTATACTAC	ACGTTGGATGTCCTCA GGAATCTAAGGCCG	TACTACACATG GAGGTCCGA	ACT	G
CCDC86	rs7167	ACGTTGGATGTTCCTC AGTGAGCTGGTGAC	ACGTTGGATGGACCTA GGATGCTGAGAAAG	CTGGCAGGTG ACTCCCTC	ACT	A
CDCA1	rs1509022	ACGTTGGATGCCTGAT TTAGTGATTGTTG	ACGTTGGATGGAAGA GTTCAAGCAGCTTTC	TGTTGTAGCTC CTGAAT	CGT	A
CEP72_HUM AN	rs2458815	ACGTTGGATGAAGAG AGAACTCTTGCCAGC	ACGTTGGATGCCCTTA AGCTTCCTGGTAAA	CTTGCCAGCCA GTACACTCA	ACT	T
CGI-69	rs2011951	ACGTTGGATGACCAGC CACAAAGCTCATGC	ACGTTGGATGTATGAG CTGGTGAAGAGCTG	GTCTGGTCCTT CGGCCTGA	ACT	A

ChGn	rs6984644	ACGTTGGATGAACCG GTGGCATCCCTCAAA	ACGTTGGATGAATGGA TGCAAGCAGCTCCG	GGCTTCCCTGG GCTAGA	ACG	C
COASY	rs615942	ACGTTGGATGCACACT CTATGTTCACCTGC	ACGTTGGATGTCAAGA ACCTCAAACGTGGC	CCCGGCTCAGC CCCAGT	CGT	C
COL18A1	rs7499	ACGTTGGATGTTCCTC CGAGCCGCCGGTC	ACGTTGGATGAGAAC AGCTTCATGACTGCC	GGTCCTCTCCG GCCATC	ACG	G
COL1A2	rs1060399	ACGTTGGATGCAGAG CATTGTGCAATACAG	ACGTTGGATGCCATAA CAGGTGTAAGAGTG	AGTTTCATTAA CTCCTTCCC	ACG	C
COMT	rs4633	ACGTTGGATGTAGGTG TCAATGGCCTCCAG	ACGTTGGATGTCATGG GTGACACCAAGGAG	TCCGCATGCTG CAGCAC	ACT	T
COQ7	rs11074359	ACGTTGGATGTGTGGG ATCAAGAAAAGGAC	ACGTTGGATGAAGGG CATCAGAACTGTTGG	GTTCAATGAGT TGATGGTTA	ACG	C
CPXM2	rs10794567	ACGTTGGATGGTGTGT TCCTGCGTCTTCCA	ACGTTGGATGAATCCC TTTTGTGCTGGGCG	CGCACCAGGT CGTAGGG	ACT	A
CRYAA	rs872331	ACGTTGGATGAAAAA CTGGTCGAACAGCCG	ACGTTGGATGCCCGTG GTACCAAAGCTGAA	GGGTGCTGGA TGGTCAC	ACT	T
CRYZ	rs17459	ACGTTGGATGGCACAC TCATGCCAAATGAC	ACGTTGGATGCTCAAA TCAAAGAGTCATC	GCCAAATGAC AACTAACATG	ACG	C
CTSD	rs12214	ACGTTGGATGAACAA AACAGCAAGTCGGGC	ACGTTGGATGACAGA AACAGAGGAGAGTCC	TGTGTGGGAG GGGCCGC	ACG	C
CYYR1	rs2830239	ACGTTGGATGGTCTCT ACCAAGAGCTTCAC	ACGTTGGATGGTTTGA GGTCCTGATGGAAG	TACTAAAACC AGAGTACCC	ACT	A
DCHS1	rs997263	ACGTTGGATGATCTCC CCCATTCCGCATTG	ACGTTGGATGGTGAAG TCAAAGCTTGAGGG	TTCCGCATTGA TGCCCA	ACG	G
DDAH1	rs233112	ACGTTGGATGGGAAA CGGTACTTTCCAATG	ACGTTGGATGGCAATA CAGGAAAGGTGGAG	GTAACTGATCC AGGCAC	ACT	T
DDIT4L	rs1053227	ACGTTGGATGAGTTGT CACCACTGACTGGG	ACGTTGGATGGCATAA CAGGCAAGAGAGTC	GTAGAACTGTT GTCCTTTT	ACG	G
DHCR7	rs1790345	ACGTTGGATGCGCGTT ATCCATGTATTGCC	ACGTTGGATGACGCTA CCAAGGAGAAACGG	TCCCCTGCCTG CTCTGCAGT	ACG	G
DISC1	rs821616	ACGTTGGATGTCTTCT ACAGACAGGCTTCC	ACGTTGGATGGAAGCT GACTTGGAAGCTTG	TTCCTGGAGCT GTAGGC	CGT	A
DLK1	rs1802710	ACGTTGGATGGCAGGT CTTGTCGATGAAGC	ACGTTGGATGCAACCC ATGCGAGAACGAC	CAGCGGAAGT CGCCCCC	ACT	T
DOCK5	rs2271108	ACGTTGGATGAGCTCA TACGGTTCTCCTTG	ACGTTGGATGTCAGCA GCCTCTTAGAGAAC	TGCATGATGAT GGTTCTATA	ACT	T
EDNRA	rs5333	ACGTTGGATGTTTGCT CTTTGCTGGTTCCC	ACGTTGGATGATTCAC ATCGGTTCTTGTCC	CTTTGCTGGTT CCCTCTTCA	ACG	C
EMILIN3	rs6072352	ACGTTGGATGTCAAGC ATGACATCCTTGCC	ACGTTGGATGAGCCAC CAAAACCCATGCAC	TGTCCGGAACT GTCCAGTTT	ACT	C
ERCC5	rs1047768	ACGTTGGATGGCAGA GCCGATGAAACAAAG	ACGTTGGATGCAAGCA CTTAAAGGAGTCCG	GGATTTTCTAT TGAGTTCCC	ACT	T
FGB	rs6056	ACGTTGGATGAGCAC ACGAAGGTTAGTTGG	ACGTTGGATGCTCAGA ACTGGAAAAGCACC	CGAAGGTTAG TTGGGATATT	ACT	T
FMO4	rs1042772	ACGTTGGATGAGTCTT TGGCGAGGATGAAC	ACGTTGGATGACTTGG AGATAGAATTAGC	GATTGTTACAA GGGTTACAC	CGT	C
FMOD	rs4605	ACGTTGGATGGCTTTT GTGGATTCCAGGTC	ACGTTGGATGTTTGAA GCACCTTCCCTGAG	GGTCTGGAGC CAAGAAC	ACT	C
FOSL2	rs7562	ACGTTGGATGGCTCAG TGCCTTTTGGTTTC	ACGTTGGATGACCAAG AGGGAGTGAAACTG	CCCTCGACTTG ACCCTT	ACG	C
GAPDH	rs1803622	ACGTTGGATGACACTC AGTCCCCCACCACA	ACGTTGGATGTAGGCC CCTCCCCTCTTCAA	CTCCCCTCCTC ACAGTT	ACT	T
GATM	rs1145086	ACGTTGGATGGCCTGT TCAGTCCAAGTAGG	ACGTTGGATGTAATGC CAATTCCCTGGGAG	TCAGTCCAAGT AGGACTGTA	ACT	A
GNAI3	rs2301230	ACGTTGGATGAAAGG CACTGCCCAAACTAC	ACGTTGGATGATTGAA TGCTCTATTGGGGG	TCAAAATGCAT TCTCTTCAT	ACG	G
GPR158	rs10828833	ACGTTGGATGTGCCTC TTCTGCTCTAAGTG	ACGTTGGATGTCTTCC TGGAGATGCTACAC	GGCCTAGGAA AGAAGAG	ACT	A
GRIA1	rs707176	ACGTTGGATGTGAAA GAGCATCCGGTATCC	ACGTTGGATGACAGCT GCTGAGAAGAACTG	TCCTCTGTGGT TGTCAA	ACT	T
GUSB	rs9530	ACGTTGGATGATGAA ACCAGGTATCCCCAC	ACGTTGGATGAGGTGG TATCAGTCTTGCTC	ACAATGTTTGG AAAACAGCC	ACG	C
HCA112	rs9088	ACGTTGGATGTGGTTG GGAACATTCTCCAG	ACGTTGGATGTCTTGG GTGTCTGGATTCTG	GTACAGCCAC AGAGGGG	ACG	G
HCLS1	rs1128163	ACGTTGGATGAACCA GATGAGGTTCTGGAC	ACGTTGGATGTCTGCC CCAAGCCCTTAATG	CTGGACAAAC TTCCCTC	ACG	C

HES6	rs9776	ACGTTGGATGCACTAG TGCCCAGCACCATT	ACGTTGGATGAGCGCT CTCCCAGGTTCAC	GCACAGGGCT GGTGAAG	ACT	C
HEY1	rs1046472	ACGTTGGATGAAGTG ATCATGGTGTGCGAG	ACGTTGGATGGCAGTC CCAGGAAAATTAGG	ATTGAGTAGTA CAGTGGAA	CGT	G
HIST1H1C	rs10425	ACGTTGGATGAGAAG AAGGCGGCCAAAAAG	ACGTTGGATGAGCCTT GGTGATGAGCTCTG	ACGCCTCGTAA GGCGTC	ACG	G
HK2	rs3821305	ACGTTGGATGAACCA AAGCCAGGAGTTGAC	ACGTTGGATGGCTTGT GGATAACATGTGCG	GCAGTTGGGC CAGCTGT	ACG	C
HLA-DPA1	rs7905	ACGTTGGATGAGTGGC TGACTGAATTGCTG	ACGTTGGATGCTTTAC TAAGGAATGACTGC	TCAAGCTCTGT CCTTATCC	ACT	T
IGF2	rs680	ACGTTGGATGAAATTC CCGTGAGAAGGGAG	ACGTTGGATGGTCCCT GAACCAGCAAAGAG	ACCTGTGATTT CTGGGG	ACG	G
IGFBP1	rs4619	ACGTTGGATGTGGGTC TCCAGAGATCAGGG	ACGTTGGATGTCATCT GGTTTCAGTTTTG	GACCCCAACT GCCAGAT	ACT	A
IL15	rs1057972	ACGTTGGATGCAATGA GAGCCAGTAGTCAG	ACGTTGGATGTCAACA GCTATGCTGGTAGG	AGTCAGTGGTT CCACAC	CGT	A
IL1RN	rs315951	ACGTTGGATGACTGAG GACCAGCCATTGAG	ACGTTGGATGGAGGCT GGTCAGTTGAAGAG	CTCAGAAGGC GTCACAA	ACT	C
ILK	rs1043388	ACGTTGGATGCAAGG CAGACATCAATGCAG	ACGTTGGATGAAAAC AGGCATAGTGCAGGG	ATCAATGCAGT GAATGAACA	ACG	C
IMPACT	rs1053474	ACGTTGGATGATGAA GCGGATGCTGTTTGG	ACGTTGGATGGTGGAA TCAGAGGGATTATG	GGAGGACAGG AAAATTTATC	ACT	A
INPP5F	rs1063224	ACGTTGGATGATAAG ATGTCGAGGATGCCC	ACGTTGGATGAACACC CATATTGGCTACTG	AAATTAAAAC CTGACTGACT	ACT	A
IRS2	rs4773092	ACGTTGGATGTGCTGC TACAGCTCCTTGC	ACGTTGGATGAGCTCA TGAGCACGTACTGG	AAGGCCCCCT ACACCTG	ACG	G
KHK	rs1131375	ACGTTGGATGGATACA GAGTGTTGCTGTCC	ACGTTGGATGTTATAA GGCCTTACCCACCC	ACACATATTGG AATTGGGGC	ACG	C
KIAA0523	rs3744725	ACGTTGGATGTTGCTT TCAAGGATCTAGGC	ACGTTGGATGCTCAGG AGTTTCCACACTGC	TCTGCCCTGGC TATCCC	ACT	T
KIAA1571	rs7582864	ACGTTGGATGTACATC ATCCTGATGTTTC	ACGTTGGATGGGTTTT CATTCCCAAGGTTC	CCTGATGTTTC TGTCCA	ACT	A
KRT6E	rs2568	ACGTTGGATGACAGTC CTCAGGCCCTTCTC	ACGTTGGATGAGAGC AGGGAAGACTAGAGG	TGGCTGCAGA GCCGTCT	ACG	G
LASS4	rs36260	ACGTTGGATGATCTGT CCAGTTGCCCAGAG	ACGTTGGATGGAAGG CTGATGATCTGTCTC	CCCAGAGGGA AGAAGGG	ACG	G
LCP1	rs11342	ACGTTGGATGCTAGGT GTGTGTGTATGTG	ACGTTGGATGTTTCTG CCACCCCTAAACTC	GCTATTGATTG GCACATATT	ACT	T
LEMD2	rs2296744	ACGTTGGATGTCTGAC TCAGAGCGATAAGC	ACGTTGGATGTCTGGA GTGGGCTGTCCTG	CCGGGCGGGG ACTTGTTCC	ACG	G
MBP	rs9199	ACGTTGGATGGGTCAT CTGCTCTAATTAGG	ACGTTGGATGGGAATC CTGTCCTCAGCTTC	TGGGATTAAA GTTTTAAGGC	ACT	A
MFGE8	rs10859	ACGTTGGATGTCACCC CTCCCTCCCTCTTT	ACGTTGGATGGAAAA CAGGACAGTGAGGAC	CCTCCCTCTTT CCCACC	ACT	G
MRPS34	rs1076695	ACGTTGGATGAGCTGC AAGAGGCGGCTCTC	ACGTTGGATGTGGACT ACGAGACCTTGACG	AGAGGCGGCT CTCGCGGCG	ACT	C
MTMR3	rs41171	ACGTTGGATGGATGG AAGACCAAGGGTTGC	ACGTTGGATGAGAAA CAAGCCATCCCTCAG	TGCCGAGCTGC CTGCTGTCA	ACT	C
MYH7B	rs2425009	ACGTTGGATGCGCTGC CGGAAGTCGGTGT	ACGTTGGATGTACACC AGCTGCGCTGCAAT	GAAGTCGGTG TAGAGCA	ACT	T
NEDD9	rs1050775	ACGTTGGATGTCATGA ACTCCAGCAACCAG	ACGTTGGATGTAATGG AGGGCGGCCATCTT	CTCTGCGAGCA GCTCAAGAC	ACG	G
NR3C2	rs5534	ACGTTGGATGGAACCT CTTTCCAAGATCAG	ACGTTGGATGTTCCAT GCACTCTCTCTCTG	TTCCAAGATCA GAAGGGAA	ACG	C
NUDCD1	rs1548082	ACGTTGGATGTAGGTA CAGCCTGTAGCAAC	ACGTTGGATGGAAAG ATGTACAATTCAGTG	AGCCTGTAGC AACTAATCAA	CGT	C
NXPH1	rs3779355	ACGTTGGATGATGTGA GAGCATGTCTTCAG	ACGTTGGATGATTTCT GCCCAGTCAGCTTC	GAGCATGTCTT CAGGTGTGA	CGT	G
OSBP2	rs2301816	ACGTTGGATGTCTCAG ATGGGATCTTGAGG	ACGTTGGATGAATGAC CTCATCGCCAAGC	AGGGAGCGCT GGAGTGC	ACT	T
PEG10	rs13073	ACGTTGGATGAGAGG AGTCTAACGGTGAAG	ACGTTGGATGATTATA CTAAAAATTTTAAG	CTTTCACATGT TCATTATCATC	ACG	C
PEG3	rs1860565	ACGTTGGATGAGAAT GGACTGAGTGAGGTG	ACGTTGGATGACGACA ACAACAGTGACG	TTCCGGGTCAT GTCGTCGTC	ACT	A
PERLD1	rs2952151	ACGTTGGATGCTGAAA GGGAGTATGGTGAG	ACGTTGGATGTGGGAT TGGGAAGGAGTTTC	GAACCTGGCT AGGGCAA	ACG	C
PHF11	rs1046295	ACGTTGGATGCTGGTT TTGAGCCTGATTGC	ACGTTGGATGACATCA ATATCATCCCTGTC	ACTCTTGGCTT AGGAAA	ACG	G

PHLDA2	rs13390	ACGTTGGATGTCTTCC ATAGCTGGAAGAGG	ACGTTGGATGACGACA TGAAATCCCCCGAC	GTCGCTGCGCT TCTCCA	ACT	A
PIK3R1	rs3756668	ACGTTGGATGAAAAC GACAAATGCGGTGGG	ACGTTGGATGCCATAT GACTTGAGTTACACC	AAAGGCATTG GGCAGAA	ACG	G
PLB1	rs2272387	ACGTTGGATGAGAGC CCTTACCTCTACACC	ACGTTGGATGACAGCC CAGTAGAGCACCTC	CCCTGCGGAA CAGCCGA	ACG	C
PLCL1	rs1064213	ACGTTGGATGAGTGCC ATGAGGATCGATTC	ACGTTGGATGGACCCG GAGTCTGAAAATTC	ATTGTGGCTGT CAGATT	ACT	A
POGK	rs10918585	ACGTTGGATGTCCATA GAAGGACACCAGAG	ACGTTGGATGGGTTAT TTCCTGGAGAAAAG	CCCATATGATG AGTCCCT	ACG	G
PON2	rs6954345	ACGTTGGATGCTTGGA GAACAGACCCATTG	ACGTTGGATGGTTCTC CGCATCCAGAACA	TCACTGTAGGC TTCTCA	ACT	C
PPP1CB	rs7475	ACGTTGGATGAGAACT TTTGCACTGTGCTG	ACGTTGGATGGAGAA ACCTGTTAACTTAC	AAGTAATACA ATGTGCCAAT	ACG	G
PRDM8	rs12780	ACGTTGGATGGAAGG ACTCATTGCAGATGG	ACGTTGGATGAAAGG AGTATGCGATGGAGC	AGTTTCTCCTC TCGCCG	ACT	G
PRKAR2B	rs257376	ACGTTGGATGGAACAT CGCTACCTATGAAG	ACGTTGGATGTTGCTT CATGCAGTGGGTTC	GTTGCCCTGTT TGGAAC	ACT	A
PSMB6	rs3169950	ACGTTGGATGAAACTT CTCGGCTTTCCCAG	ACGTTGGATGAAAGAT GGCGGCTACCTTAC	CTTTCCCAGTC TGGAGTGAA	ACG	G
PTGFR	rs899	ACGTTGGATGGAACCT GCTCTCAACAGAAT	ACGTTGGATGGACTAT CTGCAGCTAGTGTG	TCTGATGTCTG TGTGTATA	ACG	G
PTPRB	rs2465811	ACGTTGGATGGAAGA AGGAGTCCACCAAAC	ACGTTGGATGTTACCT CCTGCTAGGTTTGG	TCCACCAAAC ATGCAAGC	ACG	C
RAFTLIN	rs842424	ACGTTGGATGACCATC CAGTGCCTGTCCTC	ACGTTGGATGTTGGTT TGTAGCCCAGGGTG	GGGAGCATTC GGGTCAA	CGT	A
RAPGEF5	rs3779069	ACGTTGGATGTACCTG TTTGTGGTTCTCGG	ACGTTGGATGTACACA TGACCTGCATTCAC	TTCTCGGTCTA GGGAAA	ACG	C
RARRES1	rs2307064	ACGTTGGATGGGAAA TGTTCTGCTCGAGTG	ACGTTGGATGCTTTTT CTCGATGAGCCGTG	CAGAAACCCA GACCAAC	ACG	G
RASGRF1	rs11855231	ACGTTGGATGCAGTAA CTTGATTGCTTCAG	ACGTTGGATGATATGT CAGCTACGGCCACC	TCAGAAATCAT CCGCAA	ACG	G
RFXDC2	rs3803459	ACGTTGGATGTCAGCA GCTACCATAGAAGG	ACGTTGGATGGTGGAA CCTTTTGGTCCTTC	ACCATAGAAG GGCAGAA	ACT	T
SACS	rs4143768	ACGTTGGATGGGATCT CTAAGTTCTGCTTC	ACGTTGGATGCACATT TTCAGAGAAACAG	AGTTCTGCTTC AATTTCTTG	ACT	A
SGCD	rs7724969	ACGTTGGATGAGAAG CAGTTTTCTTTCTGG	ACGTTGGATGACACAT GTGTCAAGCACAGG	TTTTCTTTCTG GTTGTGAGT	CGT	T
SHRM	rs3733242	ACGTTGGATGCAGAA AGAGGCCAAGAAGAG	ACGTTGGATGAATGGC TTTGCAGCCTGAAC	CCAAGAAGAG ATGCTGC	ACG	C
SILV	rs1052165	ACGTTGGATGAATCAG CCTCTGACCTTTGC	ACGTTGGATGGTAGGA GAGGTCAGCTTCAG	CTCCAGCTCCA TGACCC	ACG	C
SLC22A2	rs694812	ACGTTGGATGCCCAAT CTAAATATGGAAGG	ACGTTGGATGCTAGGT TAGACTTAAAGTG	ATATGGAAGG ACCTCATGAT	ACG	C
SLC22A3	rs2076828	ACGTTGGATGCTGCTG TACAATCTGAGGAC	ACGTTGGATGAGGAG AGCCAAAAATGTCCC	AATCTGAGGA CTTGGCT	ACT	C
SLC27A2	rs1648348	ACGTTGGATGTTTCTC CAGGAACGCCCGCA	ACGTTGGATGAGGCTA CTTCTTGAAGGTGG	GCGCGCCGGC CGCCGCT	ACT	A
SLC2A1	rs2229682	ACGTTGGATGCTGCTG CTGAGCATCATCTT	ACGTTGGATGAGAAG GGCAGCACGATGCAC	AGCATCATCTT CATCCC	ACT	A
SLC40A1	rs2304704	ACGTTGGATGGGGTTT TCTGGTAAACCTTC	ACGTTGGATGGGAACT TGGTATCCATGTGC	TAAACCTTCCA GAGCAG	ACT	A
SMARCA3	rs2119342	ACGTTGGATGACTGCA TCACAAATCGGAGG	ACGTTGGATGGACCTT TTATACTTTTGACC	CAAATCGGAG GCTTTGG	ACG	C
SNX19	rs3751037	ACGTTGGATGGGTGG GATGTGTGAAGAAAG	ACGTTGGATGGTCTTC ACATAAGAACAGGG	CATTTCCTACA GCCAAAT	ACT	G
SPARCL1	rs9933	ACGTTGGATGAAAAG AGGCTTTTGGCTGGG	ACGTTGGATGCAGGAT ACACATACATGTGG	GGGGACCATC CCATTGA	ACG	C
SRP14	rs16924528	ACGTTGGATGTGATAG CTTGCTCTTCACAG	ACGTTGGATGAATTGG TCTGTGCTGCATGG	CTCTTCACAGA GATGTCTAC	ACT	A
SSNA1	rs3087779	ACGTTGGATGTGTCAC TGATGCCAGTGGAG	ACGTTGGATGTACCTT CCCTCACTGTCTC	CTGGCACCTGC CCCTCT	ACG	G
ST8SIA4	rs1428439	ACGTTGGATGTTCGAG GACCAGAGGGAGC	ACGTTGGATGAAAATG CGAGGAGAGCTTGG	GCCACAGAAG ACCCCAG	ACG	G
STOX1	rs10509305	ACGTTGGATGGGGAA TGACTTTTCCTCAAG	ACGTTGGATGCTGTCC CTTTATGGAAAGC	CATTTTCTCCA CCATACAC	ACT	A
TCF20	rs2070116	ACGTTGGATGACCGA AGACTCTGTCACAAC	ACGTTGGATGCTAGCA GCACTGAAAGCAAG	GGTCCCTGCAG CATAAAGGA	ACG	C

TF	rs8649	ACGTTGGATGAGCCGT AGTATCCCTCTTTG	ACGTTGGATGTGTAAG CTGTGTATGGGCTC	TTGTTGTTGGG TTCACA	ACT	G
TGFBI	rs1054124	ACGTTGGATGTATGTG TGCTGAAGCCATCG	ACGTTGGATGTTGATA GTGAGCATGTCCCC	ATCGTTGCGGG GCTGTCTGT	ACT	A
TRMT12	rs3812475	ACGTTGGATGGCCTCT GGAGATATTCTCTG	ACGTTGGATGTGGCTG TTGTCGCAGTTGTG	GTATCGCTGGG TAAACC	ACT	T
TSPAN4	rs7091	ACGTTGGATGTGGGTT CCGCAGCCAGCAC	ACGTTGGATGAGCCCC CGGAACCCTGTTT	CCTGAAGCCA CCTGAGC	ACG	G
TSPYL4	rs2232472	ACGTTGGATGAAGTTT GGCCGCATGCGAAG	ACGTTGGATGCCCAGA AACCTGGGATATTC	CGCATGCGAA GGCTCCA	ACG	C
TSSC4	rs1057769	ACGTTGGATGTCAGCA AGGTCTCTGCTTGG	ACGTTGGATGTGGAAT GGGAGGAAGCCTG	TGGGAGGATG CTGCCCA	ACT	T
USP29	rs3764574	ACGTTGGATGGAACTC TACGAATGTATTGG	ACGTTGGATGGCTTGT CGAGGTTCATCTTC	AGTTCTGGATG TTGAGG	ACG	G
WNT2	rs2024233	ACGTTGGATGTAGGG AATGAGAGTTCCTTC	ACGTTGGATGTTGCCA GACCATTCATCTTC	GTTCCTTCTCA CTTAGAATC	ACG	G
YOD1	rs2629665	ACGTTGGATGGCACTT GGATCTTTTTCAGC	ACGTTGGATGCCTTCA CCAAGTGAAAGTGC	GGATCTTTTTC AGCATATCT	ACT	A
ZF	rs7116195	ACGTTGGATGGGGAA TAAATCCCACCTTAG	ACGTTGGATGTGGTTC TGGCCACTTAAATG	TCCCACCTTAG GAAAATAGT	ACT	A
ZNF346	rs251848	ACGTTGGATGTCTCTT TCTTTCCCTGGGAC	ACGTTGGATGTTGCTG TGGTGCTAGCCTTC	GGGACCCAGA AATCACC	ACG	G

Appendix 3: List of primers for conventional PCR

RASGRF1		
rasgr2230518F	273 bp	GTTCTTCGGACAAGGATGGA
rasgr2230518R		GAGACTTTGAGCCACGTCTTTT
rasgr1562008F	291 bp	CACCCGCAAGTTCTCCTC
rasgrf1562008R		CTGAGCGCTGAAGGGTCTT
PHACTR2		
gDNA-rs1082	Forward	GCATGCAGAATGTGCTCCTA
Universal-rs1082	Reverse	TGGGAACTTTCCATGTTCTG
Long transcript-rs1082	Forward	AAGGACAGGGTGCTTCGTTA
Exon 5-rs1082	Forward	CAGCCACAAAGGTGATGAAG
Short transcript-rs1082	Forward	GCGGGGTAGAAGGTGAGG
PHACpGF		GATTTTAGTTATGGGTTAGATTT
PHACpGR	242bp	TAAATCCCAAAAAACATTCCTCTT
ZNF331		
short-rs8100247	Reverse	GCGGGCAGATAACTTGAAAT
Z100SNPF1		GTTTTTTTTTAGTTTTATTTTTTTT
Z100SNPR1	346 bp	CATATACACACAAACCCCTACAC
Z83SNP2F		TTTTTGTATTGTGTGGAGTGG
Z83SNP2R	192 bp	AAATTATAAATTATATTATTATATATATTT
ZNF CpG83R1		TCCACACAATACAAAAACACATC
ZNF CpG83SNPF1	282 bp	GGGTTTTTTTTATAATTAGGGTTGG
ZNF CpG83SNPR1		TAATAACTCAACCTCACCACAACAC
CpG86 F	273 bp	GGGGTTGTTTATGTATTTTTTA
CpG86 R		CCTCAAAAAAAAACCAAAATATTAATTC
ZNF331		
ZNF CpG83F1	263 bp	GTTTATGGTTTTAAGGGGAGA
ZNF CpG83R1		TCCACACAATACAAAAACACATC
ZNF CpG83R2	261 bp	CACACAATACAAAAACACATCC
ZNF CpG83SNPF1	282 bp	GGGTTTTTTTTATAATTAGGGTTGG
ZNF CpG83SNPR1		TAATAACTCAACCTCACCACAACAC

rs8109631-universal R		GCCTTCCCACAGTCCTTACA
rs8109631-gDNA F		TTCGTTTGTACTTGTGTCCCTCT
rs8109631-exon12 F		CGTAGCCATAGACTTTTCTCAGG
rs8100247-Universal F		AGGTGTGTGCGACCATGT
rs8100247-gDNA R		GTGAGTTTTTCAGAGAAATGCTG
rs8100247-exon2 R		CGGGATGCTTTTCCTGAAG
2073214inexF	228 bp	AGAACACTGAAAACCACTCTGAA
2073214inexR		TCGTGTTTCGGCTTGCTG
M13 forward	Q5601 Promega	CGCCAGGGTTTTCCCAGTCACGAC
M13 reverse	Q5421 Promega	TCACACAGGAAACAGCTATGAC
miRNA F	312bp	GGAAAATCAAATTCAGGGAG
miRNA R		GAGTATTAGGGGCAGAGAAT
Restriction primers		
ZNF 45 F1		TAGAGGTTTTTGGGGTATGGTTT
ZNF 45 R1		ATATTTCCCAACATACAACTCTACAC
ZNF 45 F2		AGAGGTTTTTGGGGTATGGTTT
ZNF 45 R2		CAATATTTCCCAACATACAACTCTACA
ZNF 100 F1		TATTTTTTAGAGATTATTTAGGAGTA
ZNF 100 R1		ATATACAAACACAACAAAACCC
ZNF 100 F2		GGATTATATTTTTTAGAGATTATTTAGGAG
ZNF 100 R2		CATATACAAACACAACAAAACCC